AN ECONOMIC ANALYSIS OF NURSES' EARNINGS IN GREAT BRITAIN

Stephen Morris

Ph.D.

City University

Department of Economics

May 2002

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to thank Debbie Morris and Ali McGuire for their support and guidance.

DECLARATION

ABSTRACT

In this thesis we examine the earnings of nurses in Great Britain. There are two general aims: to delineate the factors that affect nurses' earnings; and, to examine the nature and magnitude of wage differentials between nurses and other workers. The characteristics of the labour market in which nurses work is first described in detail and the process by which nurses' pay is determined is also discussed. We then provide a thorough and wide-ranging analysis of nurses' earnings. We calculate the private internal rate of return and private net present value to becoming a nurse. The calculations are made using the standard equations inputted with data from the New Earnings Survey and the British Household Panel Survey. The outcome is that on financial grounds in terms of relative earnings there is a rationale for choosing to be employed as a nurse. We also conduct an earnings function analysis. The determinants of wages for nurses and other workers are analysed using a novel 'double selectivity' model as well as the more common single selectivity approach. We also examine the nature and magnitude of the nurse/non-nurse wage differential. Utilising data from the Quarterly Labour Force Survey we find that nurses receive higher mean hourly wages than other workers. This difference is due partly to their superior individual and productive characteristics. We also find however that after controlling for these characteristics and selection bias the returns to endowments are also on average higher for nurses than other workers. The main finding of the thesis is that there are financial returns to being employed as a nurse in Great Britain. We discuss some policy implications of the analysis in terms of the bargaining strategy of the Staff Side and the Management Side in pay negotiations. We then discuss some suggestions for reducing the current nursing shortage.

PUBLICATIONS AND CONFERENCE PAPERS

At the time of submission of this thesis a paper consisting of a substantial part of Chapter 3 has been accepted for publication in *Applied Economics*. The complete reference is: Morris S. and McGuire A. The private net present value and private internal rate of return to becoming a nurse in Great Britain. *Applied Economics* forthcoming.

Earlier versions of Chapters 2, 3, 4 and 5 have been presented at various conferences, detailed as follows:

Chapter 2: Morris S. A review of earnings in Britain since 1970 with specific reference to nurses. Health Economists' Study Group, University of Sheffield, January 1997.

Chapter 3: Morris S. and McGuire A. The returns to nurse training in the UK. Health Economists' Study Group, University of Birmingham, January 1999.

Chapter 4: Morris S. and McGuire A. Wage differentials between nurses and other workers in Great Britain: an earnings function analysis with correction for occupation selection bias. International Health Economics Association, University of York, July 2001.

Chapter 5: Morris S. and McGuire A. A double selectivity model of earnings for nurses in Great Britain. Health Economists' Study Group, City University, September 2001.

INTRODUCTION

In this thesis we examine the earnings of nurses in Great Britain. This subject is of interest for a number of reasons. First because of the size of the nursing workforce. The National Health Service (NHS) in Great Britain currently employs some 415,000 whole-time equivalent qualified and unqualified nurses (Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 2001) making it the single largest workforce in Europe. The implication is that the level of nurses' earnings affects a significant proportion of the population. The second reason, which is partly a consequence of the size of the workforce, is that the government-funded NHS spends a considerable amount of resources each year training, managing and employing nurses. In 2000 the total nursing pay-bill for Great Britain was £10 billion (Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 2001). This represents approximately 1% of UK GDP, 3% of total public expenditure and 20% of NHS expenditure. Therefore nursing accounts for a large proportion of total public expenditure in Great Britain. Third, nurses' earnings are an important issue because of the recruitment and retention problems that currently persist in the profession, to which relative earnings are likely to contribute. The current vacancy rate for nursing posts is around 2% (Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 1998). While this percentage appears small when one considers the size of the nursing workforce the numbers of unfilled whole-time equivalent nursing posts indicated by these vacancy rates (at least 8,000) becomes significant in absolute terms. The Pay Review Body acknowledges this problem and in its most recent report concluded that "[r]ecruitment and retention remain equally important areas requiring concentrated effort if the situation is to improve." The main factors affecting recruitment and retention in the nursing profession, as categorised by the Pay Review Body, are quality of working life, workload, training and development, and fairness and comparability in levels of pay. We concentrate primarily on the last issue.

Thus from a policy perspective closer examination of nurses' earnings is clearly warranted. Additionally, there are academic reasons why an in-depth study of nurses' earnings is of interest. They arise from the unusual and complex nature of the market. As will be discussed the NHS nursing labour market is characterised by much intervention by the government in defining the expenditure limits within which wage and employment decisions are constrained. To complicate matters further there is an independent Pay Review Body that ultimately determines levels of pay, but whose recommendations are influenced heavily by the monopsony power of NHS employers. The upshot of this complex mechanism by which nurses' pay is determined is that ultimately it is the labour supply decisions of nurses and potential new entrants into the profession at the given wage that is the dominant feature of the labour market. The supply decision is separable into two related choices: a labour market participation decision; and, an occupation selection decision. The implication is that considering nurses' earnings in isolation without also incorporating these effects omits potentially important information on the factors affecting nurses' earnings. Therefore, a comprehensive analysis of nurses' earnings will also include an analysis of the decision by households to participate in the labour market and of the decision to be employed as a nurse.

In summary, nurses' earnings are an important issue, from both a policy and an academic viewpoint. Unfortunately as we note throughout the thesis this is a topic about which surprisingly little is known. In an attempt to remedy this situation we provide a rigorous and quantitative analysis of nurses' earnings in Great Britain. There are two general aims. The first is to delineate the factors that affect nurses' earnings. This allows us to understand for example why some nurses earn higher wages than others. This also admits an examination of nurses' relative earnings compared to workers in other occupations. The second general aim is therefore to examine the nature and magnitude of wage differentials between nurses and other workers to see whether nurses and other workers earn comparable wages when other factors are held constant. We use a variety of sophisticated statistical techniques to address these issues. The results allow us to draw some conclusions as to the relative attractiveness of a career in nursing, which may then be applied to the current recruitment and retention problems, and the bargaining strategies of nurses and employers in pay negotiations.

We begin our analysis by examining the structure of the nursing labour market and the mechanisms by which nurses' pay is determined in Great Britain. We identify the determinants of the demand and supply of nursing labour and then study their interaction. On the demand side we emphasise the crucial role of the government in setting the NHS budget and, as a consequence, defining the expenditure limits within which wage and employment decisions are constrained. Also important is the Pay Review Body that determines levels of pay, but whose recommendations are influenced heavily by the monopsony power exerted by NHS employers. On the supply side we find the supply decisions of individuals in terms of joining and leaving the profession are paramount in determining the state of the labour market. We also discuss at length in Chapter 1 the process by which nurses' pay is determined. The upshot of the discussion is that the pay levels are recommended by the Pay Review Body based on the strength of the evidence submitted from the Staff Side, the Management Side and from the wider economy. While the Pay Review Body takes a number of issues into account in its deliberations (affordability of potential pay rises, recruitment and retention, fairness and comparability, morale and motivation, and productivity and workload) the evidence suggests that in the past the issue of affordability stressed frequently by the monoponistic employers is given much prominence, though more recently it is recognised that recruitment and retention are of prime importance. The outcome is that at least up until recently the market wage rate has been set by the Pay Review Body more in line with the preferences of the Management Side, as opposed to the higher wage levels preferred by the Staff Side.

Before a thorough analysis of nurses' earnings may proceed more information is needed about how their earnings have fared in recent years. Chapter 2 considers this issue. First we examine trends in nurses' pay determination over time. We focus particularly on the role of the Pay Review Body, which since it was established in 1983 has been responsible for setting pay levels for nurses in Great Britain. We find that throughout this period there has been relatively little variability in nurses' pay and there have been comparative improvements in earnings in real terms. Also in this chapter we review nurses' actual earnings over time illustrating the relationship between events in nurses' pay history and their earnings. We also compare nurses' earnings with those of workers in other occupation groups.

Following the comprehensive picture of the nursing labour market provided in Chapters 1 and 2 we then go on to analyse in more detail in Chapter 3 the financial returns to being employed as a nurse. We examine the lifetime costs and benefits. We measure the private net present value and the private internal rate of return to being employed as a nurse. From the literature review in this chapter we find that while the number of studies measuring the attractiveness of investments in human capital in this way is massive there has to date been no comparable study of the returns to nursing in Great Britain. In considering the returns to being employed as a nurse we outline the conditions under which a career in nursing is likely to be considered attractive relative to a career in some alternative occupation. The conclusion is that on financial grounds in terms of their earnings there is a rationale for individuals to choose to be employed as nurses rather than an alternative occupation.

In Chapter 3 we analyse in some detail nurses' earnings and the returns to nursing. In Chapters 4 to 6 we develop the analysis further and look behind the earnings of nurses that are taken as given in Chapter 3 to understand *why* nurses' earnings are of the magnitude they are. We examine the factors that affect nurses' earnings and look more closely at the nature and magnitude of wage differentials between nurses and other workers. In Chapter 4 we develop a theoretical model of earnings to be used in subsequent chapters. The model is based on the work of Jacob Mincer who shows in a framework suitable for econometric estimation that two important factors driving earnings are the amount of compulsory and non-compulsory education received and years of work experience (Mincer, 1974). We supplement the Mincerian model with an examination of labour market participation and occupation selection decisions. This is relevant because as has been shown by the work of James Heckman failure to account for the self-selected nature of the decision to participate in the labour market and the decision to choose to be employed in a particular occupation leads to biased estimates of the Mincerian earnings function (Heckman, 1979). The problem is particularly relevant to the nursing labour market because within the framework by which nurses' pay is determined the supply-side decisions drive actual earnings. Following the exposition we go on to review the literature on earnings function for nurses. We find that the majority of studies to date have concentrated on the US nursing labour market and are not relevant to British nurses. It is also notable that these studies suffer frequently from selection bias problems of the kind alluded to above. There has so far been a single earnings function analysis for nurses in the NHS in Great Britain (Phillips, 1995) conducted as part of a wider analysis of nursing labour supply, though this study is now quite dated (1980).

In Chapter 5 we analyse the earnings of nurses in Great Britain using the economic model constructed in Chapter 4. We in fact estimate five statistical models that involve the estimation of wage equations for nurses and other workers with corrections for participation selection bias and occupation selection bias using the Heckman two-step procedure. The basic methodology is to include in the wage equations for nurses and other workers selection bias correction terms that control for the effect on earnings of the propensity to participate in the labour market and the propensity to be employed as a nurse. The data to which the models are applied are taken from the Quarterly Labour Force Survey. The final sample consists of 247,774 females aged 18 to 60 years of whom 8,878 are employed as nurses. The results show the factors that affect nurses' earnings. Using an algebraic method developed by Ronald Oaxaca (Oaxaca, 1973) we go on to decompose the observed earnings differential in the sample between nurses and other workers into differences in labour market endowments and differences in the returns to these endowments. We find that nurses in the sample earn on average higher wages than other workers, but that this differential is due almost exclusively to their superior human capital characteristics (higher educational attainment, greater years of experience, etc.).

An important finding in Chapter 5 is that the decision to participate in the labour market and the decision to choose to be employed as a nurse are potentially important factors affecting the earnings of nurses. The key point is that while in the previous chapter we corrected for these two sources of potential selection bias we made the corrections individually in separate models. In Chapter 6 we build on this work and construct extended earnings functions for nurses and other workers in Great Britain correcting jointly for *both* participation selection bias *and* occupation selection bias. We estimate four statistical models using a bivariate probit selection framework and a multinomial logit selection framework which treat in different ways the effects of both the participation and the occupation selection decisions. Because they correct simultaneously for two forms of selection bias these models are referred to as 'double selectivity models'. In the review section we find that these types of model are extremely rare in the literature, and there has been only a single application to the (US) nursing labour market (Botelho et al., 1998) though the model was used in a different context. As in the previous chapter we supplement the basic results with a decomposition analysis. We find again that the higher wages earned by nurses are attributable mainly to their superior human capital endowments. In addition we find that even after controlling for differences in labour market and individual productive characteristics there are financial returns to being employed as a nurse in Great Britain relative to other occupations.

Chapter 7 concludes by pulling together the findings of the first six chapters. We also discuss some policy implications of the analysis. We first relate the finding of a positive wage premium to being employed as a nurse to the process by which nurses' pay is determined as described in Chapter 1 and the bargaining strategy of the Staff Side and the Management Side in pay negotiations. We then discuss some suggestions for reducing the current nursing shortage in light of the findings of the thesis.

The two general aims are fulfilled: we identify the factors that affect nurses' earnings; and, we examine the nature and magnitude of wage differentials between nurses and other workers. We noted at the outset that surprisingly little is known about the earnings of nurses in Great Britain and this thesis seeks to remedy in some part this problem. The thesis itself makes an original contribution to the literature in three major respects. First, in Chapter 3 we conduct an analysis of the private internal rate of return and the private net present value to becoming a nurse. This is the first time such an analysis has been conducted for nurses in Great Britain. Second, in Chapters 4 and 5 we estimate earnings functions for nurses with appropriate corrections for selection bias. The methodology employed is now fairly standard, but the application to the British nursing labour market here is unique. The third original contribution is provided in Chapter 6. Here we analyse the factors affecting nurses' earnings within a double selectivity framework. We construct a novel set of four statistical models to analyse nurses' earnings simultaneously adjusting for the effects of the decision to participate in the labour market and the decision to be employed as a nurse.

CHAPTER 1

NURSES' PAY DETERMINATION IN GREAT BRITAIN

1.1 Introduction

In this thesis we conduct an economic analysis of nurses' earnings in Great Britain. [1] We begin in this chapter by examining the structure of the nursing labour market and the mechanisms by which nurses' pay is determined. This forms the basis for the analysis of the following chapters. We start by examining the demand and supply mechanisms that operate in the labour market for nurses. We analyse the unusual structure of the market in order to explain why the conventional perfectly competitive labour market model does not apply and to help understand the prevailing conditions vis-à-vis wages and employment. We then go on to examine the complicated mechanisms by which nurses' pay is determined. We first place the discussion in a historical context and then discuss the process by which nurses' pay in the NHS is determined under the current system. We focus on the role of the Pay Review Body which since its inception in 1983 has been responsible for setting pay levels for nurses in Great Britain.

The chapter consists of three main sections. In the first we consider the demand for nursing labour. We find that determining demand is fraught with difficulties and should effectively be viewed as the number of nurses employers can afford to employ according to their budget

[1] The term 'nurse' used throughout this thesis relates to qualified nurses (C grade and above) working in the nursing profession. In later chapters we also consider individuals not actually participating in the labour market who consider themselves to be nurses. We generally do not include unqualified nurses (A and B grades – nursing auxiliaries and assistants) whose education and training and conditions of pay and employment are different to their qualified counterparts. For the purposes of the thesis we will normally concentrate specifically on nurses, as opposed to midwives and health visitors, who generally perceive themselves to be separate

and the given wage rate. We then consider the supply of nursing labour in the second section. This allows a more conventional approach and we find that the effect of the numbers of joiners and leavers on the stock of labour ultimately determines the quantity of nurses employed. In the third section we describe in detail the role of the Pay Review Body in determining nurses' pay and how it recommends pay levels for nurses on the basis of the submitted evidence from nurses and their representatives, employers and their representatives and from the wider economy. We conclude that while the Pay Review Body considers a number of issues in its deliberations in recent years evidence suggests that the issue of affordability has been an important principle on which it has based its recommendations. We also find that increasingly the issues of recruitment and retention are becoming prominent. The implication is that the wage rate has in recent years been set by the Pay Review Body below the level preferred by nurses and their representatives and that it is the supply decisions of individuals in terms of joining and leaving the profession are paramount in determining the state of the labour market.

1.2. The demand for nursing labour

Most nurses in Great Britain (some 85%, Department of Health, 2001) are employed in some capacity or other – either in hospitals or in the community – by the National Health Service (NHS). The NHS is the main provider of health care in Great Britain and nursing labour is an important input into the production process. The demand for nursing labour as a factor of production is therefore a derived demand – derived from the demand for health care. This in turn is derived from the demand for health by the population, which is a function of the

occupation groups. We also focus on nurses who are employed by the NHS, which constitutes 85% of all qualified nurses (Department of Health, 2001).

consumption and investment benefits associated with improved health status. Because of these inter-relationships and because of the political context within which the NHS nursing labour market operates defining the demand for nursing labour is far from straightforward. There are five main difficulties: estimating the demand for health care; the role of the government; quantifying the contribution of nursing labour to the production of health care; the role of wages in the demand for nursing labour; and, the existence of non-competitive behaviour.

The demand for goods and services is usually measured in terms of the responsiveness of quantity demanded to price allowing for the interaction with additional variables such as income, tastes and the prices of other goods. The demand for health care does not fit neatly into this model. This is for a number of reasons. First is the fact that in the NHS health care is effectively free at the point of receipt for the majority of services – under a primarily taxation-based system the price and consumption of health care are separated. Second, given the short-run disutility associated with its consumption it is unclear whether the quantity of health care demanded will react inversely to changes in price in the way that conventional economic theory predicts. Third, the market for health care is characterised by imperfect knowledge because the majority of the population have little or no knowledge of health and health care. This solution is remedied with the introduction of principal-agent relationships between individuals in the population and health care professionals/providers. Unfortunately the asymmetric information characterised by the agency model gives rise to potential incentive problems and moral hazards may occur with the introduction of supplier-induced demand. Given these market distortions it is difficult to see how a demand function for health care in the NHS could ever be identified entirely accurately.

An equally important and problematic issue is that in a taxation-based NHS the total amount of health care consumed in any one year will be determined not on the basis of clinical need but instead by the size of the NHS budget. This will in turn be affected by the performance of the economy, the rate of taxation and the proportion of total public expenditure collected in taxes that are allocated to the service. Through a series of complex negotiations between various government departments the Treasury determines the proportion of total public expenditure allocated to the NHS. Via the public expenditure survey this is then ratified by parliament each year. In Figure 1.1 some data on NHS expenditure are provided. The growth rate of real GDP and real NHS expenditure are provided along with the proportion of total public expenditure spent on the NHS. With the exception of one year in the early 1990s when real GDP fell and real NHS expenditure increased (presumably as a result of the funding required to implement the NHS reforms) the growth rate of GDP and of NHS expenditure have generally followed very similar paths.

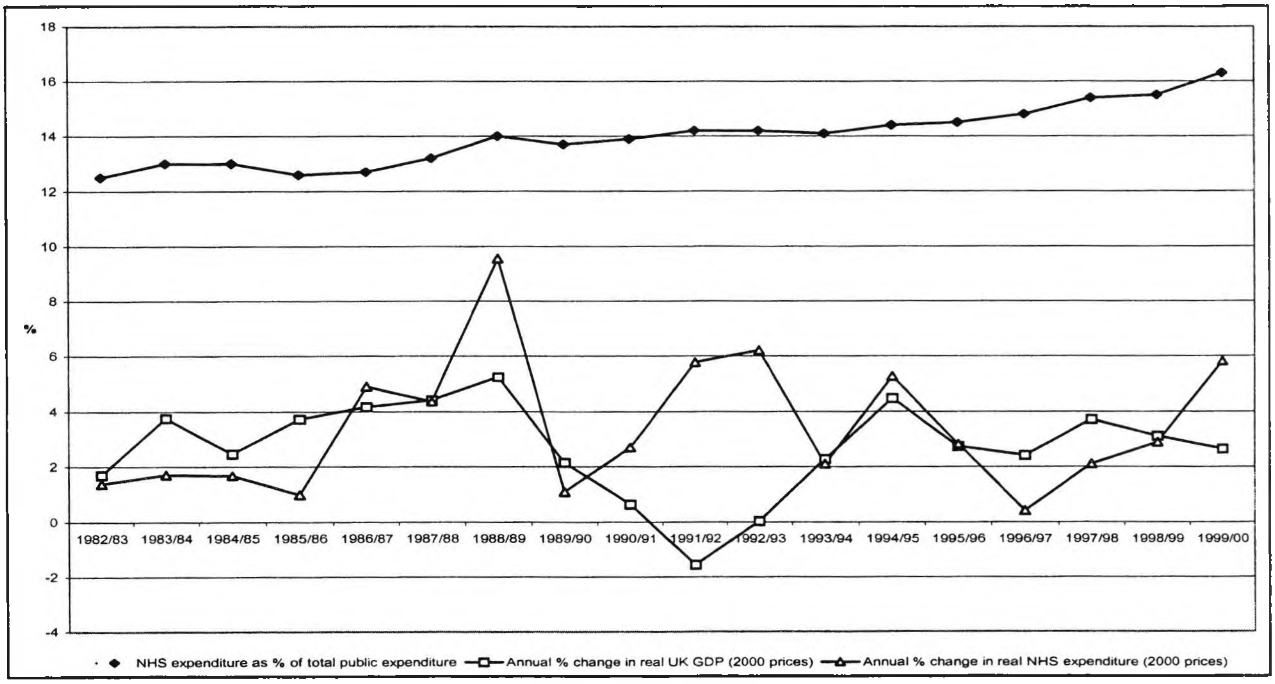


Figure 1.1. % Change in real GDP and real NHS expenditure, and % of total public expenditure spent on the NHS, 1982-2000 [source: ONS. UK National Accounts (selected years), ONS. Annual Abstract of Statistics (selected years)]

What these data suggest is that the NHS budget which influences the amount of services provided (and which in turn affects directly the demand for nursing labour) is closely linked to economic performance. It would appear that this, rather than total need, determines the consumption of health care in the NHS. Note also from Figure 1.1 that there has been a steady growth in NHS expenditure as a percentage of total public expenditure over time, rising from around 12% in 1982 to around 16% in 2000.

The demand for health care in the UK is therefore a complex issue, characterised by imperfect knowledge, asymmetric information, zero pricing at the point of receipt, and substantial government involvement. These features arise due to the nature of the good health care and as a direct consequence of a taxation-based system. Difficulties then arise in delineating the demand for nursing labour, which is derived from the demand for health care. In a perfectly competitive model the employer takes the wage rate as given and continues to hire workers up to the point at which the marginal value product (MVP) to the employer from employing an additional worker is equal to the wage rate. This in turn is affected by the additional contribution to output of employing additional workers (the marginal product of labour, MP_L) and the price of the final good. There are two problems in applying this to the field of nursing. The first is that as noted above there is not a product market for health care in the conventional sense. Because there is effectively no market price it is difficult to value the marginal product of labour.

The second problem concerns the impact of nursing staff on output. It is difficult to measure the contribution of nurses (and many other inputs for that matter) to the production of health care. Difficulties in measuring the MP_L arise first because no ideal method exists for measuring the output of the NHS. A number of measures do exist, such as average daily available and occupied beds, numbers of finished consultant episodes, numbers of discharges and deaths, numbers of admissions, and average length of stay. While these reflect in some dimensions the output of the NHS, no single measure captures accurately and comprehensively the full range and amount of the activities provided. The second difficulty is that it is hard to separate out the specific contribution of the various inputs into the production process. There are many factors that go into the provision of health care including many different types and grade of health care professional, non-labour disposable inputs such as medical supplies, and capital inputs such as equipment and machinery. On this basis it is difficult to measure and quantify the contribution to output of individual nursing staff.

A number of measures of nursing workload (called nursing workload management systems – NWMs) do exist, such as the Financial Information Project, EXCELCARE and Criteria for Care (see Jenkins-Clarke, 1992, for a review). Workload is described either in terms of the aggregation of time spent on individual activities for each patient, or as a measure of the relationship between the number of nurses working on a ward and total activity. However, it has been shown (Jenkins-Clarke, 1992) that there are substantial problems associated with accuracy, consistency, reliability and implementation of different NWMs. Also, while NWMs do quantify the workload of nurses (albeit with varying degrees of success), they do not measure MP_L or contribution to output.

According to standard economic theory one of the determinants of the demand for labour is the degree of substitutability with other factors of production. Thus the issues of skill-mix and staff substitution become important. In the context of the demand for nursing labour different combinations of health care professionals may be used to provide health care within a given setting. Policy makers in a scarcely resourced NHS are aware of the need to employ labour in a cost-effective manner. This could influence the demand for qualified nursing labour in a number of ways: the substitution of doctors for nurses; the substitution of higher grade for lower grade qualified nursing staff; and, the substitution of qualified for unqualified nurses. In terms of substituting doctors for nurses, if the substitution of higher paid staff with those who are lower paid (or vice versa) is cost-effective then this has direct implications for the demand for nursing labour. For example, a recent review of the literature suggests that between 30% and 70% of the tasks performed by doctors could be carried out by nurses (Richardson and Maynard, 1995). Since nurses' earnings are lower this suggests that a cost-effective substitution is possible. This has obvious implications for the demand for nursing labour. Unfortunately, as Richardson and Maynard (1995) note there has generally been very little evaluation of the effectiveness and cost-effectiveness of doctor/nurse and qualified/unqualified nurse substitution. Partly this is due to methodological difficulties, some which have been discussed already. It is often difficult to measure the effectiveness/output of health care. It is also difficult to identify the proportion of a particular task in health care that is attributable to an individual worker. These problems prevent valid conclusions from being drawn on the substitutability of doctors for nurses. This further demonstrates the difficulties in defining the demand for nursing labour.

In terms of substituting between different grades of nursing staff, in one of the few UK-based studies of its kind Carr-Hill et al. (1995) examine the impact of nursing grade on the quality and outcome of nursing care. Using a Quality of Patient Care Scale (QUALPACS) the authors analysed the relation between the skill-mix of a group of nurses (measured in terms of grade) and the quality of care provided based on data collected from 15 wards at 7 sites. The general finding was a positive relationship between grade and quality of care. For example, a positive correlation was found between the proportion of nursing staff on a ward on grade D or above and QUALPACS score (Pearson correlation co-efficient = 0.53, p=0.02). It was also found on average that enriching the skill-mix by one grade led to an overall increase in the quality of care. However, while providing a useful insight into the effect of nursing skill-mix on patient care the authors did not seek to measure fully the impact of nursing staff on output. Particularly, the authors were unable to tease out the specific contribution of nursing staff to the production of health care. As they pointed out: "It was also impossible to control for the input of other disciplines to patient care or patient length of stay" (Carr-Hill et al., 1995).

A further difficulty relates to the importance of wages in the demand for nursing labour. This reflects the role of the government in the nursing labour market. First, NHS nurses work in the public sector and as noted by Elliott and Duffus (1995) their wages are therefore determined within the confines of public sector expenditure constraints. Being in the public sector also means that in reality the demand for nursing labour is likely to be determined less by the interaction of MVP and wages and more by the amount of resources allocated to health care providers (employers of nursing labour) for the provision of health care services. As discussed above the size of the annual NHS budget is the result of a political decision made by the government and is decided in the broader context of expenditure across the entire public sector. Funds earmarked for the NHS are allocated by the Treasury to the Department of Health and are then passed down through the system and eventually to health care providers to meet the costs of the services they provide. Through this allocation of funds each health care provider is effectively allocated an annual budget with which it is required to provide (meet the cost of providing) all contracted health services. At any given wage rate therefore, the government effectively determines the number of nurses employed by health care providers each year. By setting the cash-limited NHS budget the government effectively sets the size of the nursing paybill which acts as a budget constraint on the number of employed workers at the given wage. The implication is that while wage rates remain a key factor affecting the demand for nursing labour they now play a slightly different role. Employers are unlikely to base employment decisions on the interaction of the wage rate and MVP but instead by the interaction of wages and a hard budget constraint. [2] The budget constraint limits the maximum number of nurses that may be feasibly employed at any give wage.

That the nursing labour market in Great Britain does not operate in a perfectly competitive environment is clear. More specifically, a further problem in determining the demand for nurses is the existence of non-competitive (monopsony) behaviour. The NHS as a whole may be thought of as a monopoly provider of health care. While the NHS reforms of the 1990s have sought to increase competition within the NHS this does not deflect from the fact that the NHS as a whole remains the primary producer in the product market. More importantly in terms of this thesis the NHS is effectively a monopsonist in the labour market, being the major employer of nursing labour (85% of qualified nursing staff are employed by the NHS, Department of Health, 2001; see section 1.3). The implication is that rather than taking the wage rate as exogenously given NHS employers attempt to influence nurses' pay. Compared to the perfectly competitive equilibrium in a monopsonistic market wages will generally be lower and there will be fewer workers employed. This is an oversimplified exposition of the market structure, however. Currently nurses' pay is determined by the Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine (hereafter called the Pay Review Body). This panel determines the levels of pay that nurses receive on

---

[2] This is predicated on the assumption that at every wage rate the quantity of nursing labour demanded shown by the MVP curve lies to the right of the number of nurses feasibly employed given the budget facing employers.

the basis of evidence from nurses and their representatives, from (monopsonistic) managers and employers and their representatives, and from the wider economy. [3] The outcome depends on recommendations of the Pay Review Body, which are affected by the strength of evidence submitted by the different sides in the negotiations, which in turn may be influenced by the monopsony power of the employers. The upshot is that the wage rate facing employers which they use to make demand-side decisions is not determined by market forces, but is instead determined in what is effectively a regulated monopsony model (see below for further discussion). This further complicates the process of defining the demand for nursing labour.

In summary, determining the demand for nursing labour in the NHS is far from straightforward. The nature of the good health care for which nursing labour is a factor of production and the political context within which the nursing labour market operates alters radically the way in which employers make demand decisions. Selecting the quantity of nurses to employ at a given wage rate based on the value of their marginal contribution to output is not feasible. A more informative way of looking at the demand for nurses in the NHS is in the context of the number of nurses employers can afford to employ at the given wage set by the Pay Review Body and the budget for nurses pay effectively set by the government.

1.3. Supply of nursing labour

We shall now examine factors that affect the supply of nursing labour in the NHS nursing labour market. Conventional theory suggests there will be a positive relationship between the

---

In simple terms the budget constraint lies to the left of the MVP curve and a lack of resources preclude employers from employing their preferred number of nurses (based on the MVP) at any given wage rate.

[3] We examine in greater detail how the Pay Review Body operates in Section 1.4.

wage rate and the quantity of nursing labour supplied by households. This positive relationship is explained in terms of workers who currently choose to work in the labour market and in terms of potential new entrants. Current workers may choose to supply more or less of their time as wages rise depending the relative magnitude of the income and substitution effects. Potential entrants view wages in alternative occupations as being relative to one another and will therefore, ceteris paribus, enter an occupation as relative wages rise, and vice versa. Essentially the supply of qualified nursing labour therefore has two components: the numbers of individuals choosing to join the profession; and, the number of leavers. In tandem these factors define the stock of nursing labour.

The numbers of joiners and leavers will be affected by individuals' labour market participation and occupation selection decisions. There are two related choices that affect the decision as to whether or not an individual will choose to be employed as a nurse in the NHS. The first choice is whether or not to work in the labour market, and the second is whether or not to choose to be employed specifically as an NHS nurse. The theory of labour supply states that individuals will choose to participate in the labour market if the wage they are offered by employers is greater than their reservation wage. The offered wage will be affected by the individual's productive characteristics such as their schooling and labour market experience. The reservation wage will be affected by property income and other variables that affect tastes for work, such as family commitments. The theory of compensating wage differentials then suggests that given the decision to participate utility-maximising individuals base their occupation selection decisions on the relative expected financial and non-financial costs and benefits of alternative occupations. These include the financial costs of training and education where this is required and the subsequent financial earnings of the individual. Non-wage factors include those experienced while undergoing education and training, and job characteristics such as general working conditions, hours worked and the level of job satisfaction. In simple terms if the expected benefits are greater than the expected costs the individual will choose to join (in the case of potential new entrants) or will remain working in (in the case of current workers) a particular occupation. If the expected costs are greater than the expected benefits then the individual will find an alternative.

Applying these ideas to the supply of nursing labour, a career in nursing has many characteristics that could be considered as unpleasant. Tasks can be dirty, tedious, and repetitive, and nurses face a substantial risk of illness from their constant exposure to sickness. Also, nurses are often required to work varying shifts and professional freedom is limited as nurses are subject to supervision from various sources, including managers and doctors. These are potentially important non-wage costs. On the other hand, there are many positive features associated with nursing such as the satisfaction derived from helping others. Also, nursing frequently offers stable employment and jobs in the profession are often readily available. There are therefore many non-wage reasons why an individual might or might not choose to become a nurse. However, there is no reason to suppose at this stage that wages are any less important to nurses than they are to individuals in other occupations.

As with demand, another important feature of the supply of nursing labour is the role of the government. While theoretically individuals will self-select occupations that maximise their utility in reality selecting an occupation is not necessarily a completely free choice. Entry into further and higher education is limited; both directly in terms of quotas of student numbers and indirectly in terms of financial constraints imposed on students. The number of available places is limited by the willingness of the government to fund them. Therefore just as the demand for nursing labour is influenced directly by government's funding decisions, so too is the number of entrants into the nursing profession. On the one hand if the number of suitable applicants is greater than the number of places funded by the government (demand for places is greater than supply) then it is the government's funding decisions that are dictating entry into the profession. On the other hand, if there are unfilled places (supply is greater than demand) then it is the individual labour supply decisions of potential new entrants that are more important.

Given this background information on the nature of nursing labour supply it is now useful to examine the actual state of play. The number of whole-time equivalent nurses working in hospitals in England across the period 1951 to 1999 are presented in Figure 1.2. This figure contains three prominent trends. First, the total numbers of staff (qualified, unqualified and learners) rose dramatically across the 1950s, 1960s and early 1970s with the expansion of the NHS and increased health care spending. Second, in the late 1970s and 1980s the growth in numbers stopped and the number of nurses remained generally constant with only small fluctuations. Third, since the late 1980s the number of nurses has declined somewhat.

The most striking trend shown in Figure 1.2 is the termination in the growth rate in the late 1970s. The explanation for this is the general economic slowdown that occurred in the UK in the mid-1970s. The effect was that the government reduced its public expenditure plans, including expenditure on the NHS. Two reports published by the then Department of Health and Social Services (DHSS) in 1976 recommended changes to the allocation of NHS resources as a result of the economic slowdown (DHSS, 1976a, DHSS, 1976b). Both announced that resources would henceforth be distributed in a systematic and transparent way related to need and, importantly, both stressed that the NHS would expand in the future at a much slower rate compared to previous years. The upshot was that growth in NHS expenditure proceeded at a much slower rate, introducing knock-on effects to the growth in the nursing labour force and resulting in the decline in the growth rate shown in Figure 1.2.

The decline in numbers in recent years is explained at least in part by changes to the nurse education system in the late 1980s with the introduction of Project 2000. Under the new system individuals training to be qualified nurses are counted as students at universities rather than as NHS employees. So since 1990 the numbers in Figure 1.2 exclude trainee nurses on Project 2000 courses.

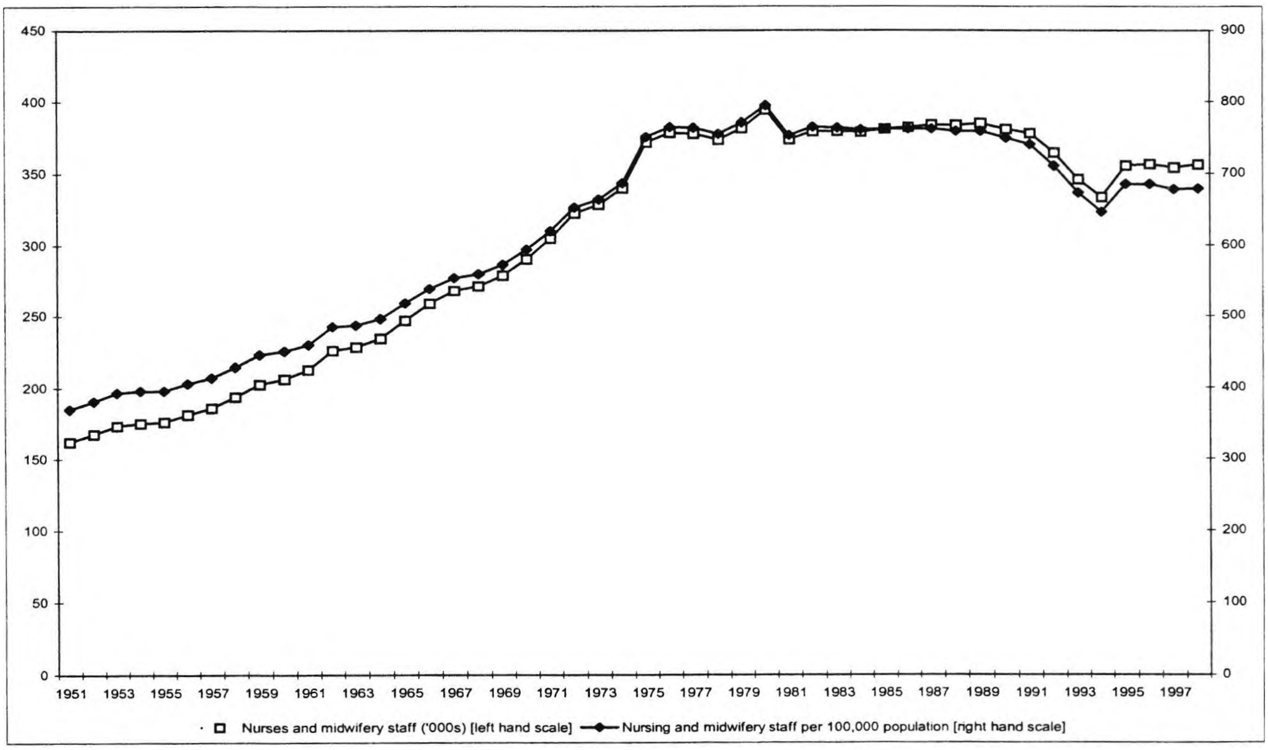


Figure 1.2. Whole-time equivalent NHS hospitals' nurses and midwifery staff, England, 1951 – 1999. [source: ONS. Health and Personal Social Services Statistics for England (selected years), ONS. Annual Abstract of Statistics (selected years)]

More information on the composition of the NHS nursing workforce is presented in Table 1.1. Based on the NHS Hospital and Community Health Services Non-Medical Workforce Census conducted by the Department of Health this shows (footnote 6) that the addition of some 30,000-40,000 individuals nursing students excluded from Figure 1.2 in the 1990s brings the total nursing workforce more in line with the constant level of the 1980s.

	1995	1996	1997	1998	1999	2000
Qualified [1,2]	246,820	248,070	246,010	247,240	250,650	256,280
Acute, elderly & general [1,2,3]	131,350	131,000	130,460	131,270	133,980	138,120
Paediatric [1,2,3]	11,300	12,590	12,590	13,080	13,380	13,640
Maternity [1,2,3]	23,090	23,190	22,780	23,060	22,920	22,780
Psychiatry [1,2,3]	34,990	35,450	35,290	34,620	34,970	35,800
Learning disabilities [1,2,3]	11,310	10,720	9,880	9,330	8,780	8,400
Community services [1,2,3]	33,040	34,400	34,420	35,300	36,060	36,870
Education staff [1,2,3]	1,750	730	580	570	560	660
Unqualified [1,2]	82,910	83,650	84,020	84,520	87,440	89,830
Learners [1,2]	4,580	2,670	2,250	2,080	1,880	1,970
Unclassifiable [1,2,4]	710	940	590	430	490	0
Total [1,2,5,6]	330,440	332,660	330,620	332,200	338,580	346,180

[1] Figures represent whole time equivalents in England as at 30th September each year.
[2] Includes nurses, midwives and health visitors.
[3] Qualified staff only.
[4] Staff for whom it was not possible to determine a specific level of qualification.
[5] Totals may not equal the sum of component parts due to rounding and the inclusion of unclassifiable staff.
[6] These figures exclude students on training courses leading to a first qualification as a nurse or midwife; there were around 34,000 in September 1996, 36,000 in September 1997, 38,000 in September 1998, 42,000 in September 1999, and 41,000 in September 2000

Table 1.1. Nursing, midwifery and health visiting staff by type and area of work in England, 1995-2000 [source: Department of Health. NHS Hospital and Community Health Services Non-Medical Workforce Census]

In terms of Figure 1.2 it would therefore appear that the decline in numbers in the 1990s is not nearly as substantial as Figure 1.2 implies. However, an important feature of supply is that while the constancy of the total numbers in recent years as shown in Table 1.1 suggests stability in the supply of nursing labour in fact this masks the true picture of continual inflows and outflows from the profession. There is a continuous flow of individuals into and out of nursing. The number of qualified nurses may be viewed as the stock of qualified nursing labour. This stock will remain constant if the numbers of joiners and leavers is equal. It may also increase or decrease over time. Decreases for example might occur because the flow of individuals into the profession declines or because the outflow of nurses from the profession grows. Table 1.2 shows recent (2000) data on joiners and leavers to the NHS nursing workforce. In 2000 there were 23,623 'qualified nursing staff joiners' representing 14.7% of the total staff in post who came from a number of sources, the most prominent being transfers from within the NHS and the newly qualified. By contrast there were 21,699 leavers (13.5% of total staff in post), who left for a variety of reasons the most important being to transfer to other NHS units, to retire or to work in the private sector. This table also serves to highlight a number of other important characteristics of the structure of nursing labour supply. First, in terms of the flow in to the labour market while the newly qualified comprise the bulk of the new entrants into qualified nursing workforce re-entrants are also important – individuals who left the nursing profession and then chose to rejoin. Also important and included in the 'Other' class are nurses from overseas who are recruited to work in the NHS. In terms of outflow note that a number of nurses leave the NHS every year to undertake non-NHS health care employment in the private sector.

	Qualified staff [1]	Unqualified staff	All staff [2]
Joiners			
Newly qualified	4,161 (2.6) [17.6]	93 (0.1) [1.0]	4,875 (1.8) [13.3]
Transfers within NHS	8,535 (5.3) [36.1]	1,213 (1.6) [13.1]	11,454 (4.2) [31.3]
Re-entrants	1,969 (1.2) [8.3]	391 (0.5) [4.2]	2,685 (1.0) [7.3]
Other	3,673 (2.2) [15.6]	3,185 (4.3) [34.3]	7,350 (2.7) [20.0]
Don't know	5,287 (3.3) [22.4]	4,410 (5.9) [47.5]	10,246 (3.7) [28.1]
Total joining	23,623 (14.7) [100.0]	9,291 (12.4) [100.0]	36,611 (13.4) [100.0]
Leavers			
Retirement	1,383 (0.1) [6.4]	743 (1.0) [8.4]	2,725 (1.0) [7.9]
Transfer to NHS units	7,161 (4.4) [33.0]	827 (1.1) [9.4]	9,269 (3.4) [26.9]
To non-NHS health units	1,106 (0.1) [5.1]	374 (0.5) [4.2]	1,599 (0.6) [4.6]
To other employment	844 (0.1) [3.9]	493 (0.7) [5.7]	1,460 (0.5) [4.2]
Other	4,734 (2.9) [21.8]	2,218 (3.0) [25.1]	7,270 (2.8) [21.1]
Don't know	6,471 (4.0) [29.8]	4,168 (5.6) [47.2]	11,725 (4.3) [34.0]
Total leaving	21,699 (13.5) [100.0]	8,824 (11.8) [100.0]	34,498 (12.6) [100.0]

[1] Does not include Midwives, Health Visitors, District Nurses, managers and education and nursery nurses.
[2] Includes Midwives, Health Visitors, District Nurses, Nurse managers and education and nursery nurses as well as qualified and unqualified nursing staff.

Table 1.2. Nurse joiners and leavers in NHS Trusts in Great Britain in the year to March 31st 1999: whole time equivalents (as % of staff in post) [as % of total joining or leaving] [source: Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine]

Inherent in the occupation selection decision to be employed as a nurse therefore is the choice to be employed by the private sector (in private nursing homes, hospitals or clinics) or the NHS.

More detail concerning the magnitude of the private sector nursing workforce is presented in Table 1.3. Note that the number of qualified private sector nurses is around one fifth the size of the qualified NHS workforce. Put another way, the NHS employs roughly 85% of all qualified nursing staff. Note also from Table 1.3 that the majority of private sector nurses (84%) work in nursing homes rather than hospitals or clinics. Private sector nurses therefore play a very different role in the provision of health care than NHS nurses, with different job specifications and job characteristics.

	1994-95	1995-96	1996-97	1997-98 [2]	1998-99 [3]	1999-2000
General nursing homes [1]	36,021	36,021	35,516	.	38,961	36,187
Mental nursing homes [1]	6,227	7,152	8,055	.	7,305	6,941
Hospitals and clinics [1]	8,037	7,635	7,655	.	8,562	8,035
Total [1]	50,465	50,808	51,226	.	54,828	51,563

[1] Figures represent whole-time equivalents in England.
[2] Figures are not available for 1997-98.
[3] A different estimation method was used in 1998-99 and therefore the figures presented may not be comparable with those presented in other years.

Table 1.3. Qualified nursing staff in England working in the private sector, by premise type, 1994-2000 [source: Department of Health. Community Care Statistics: Private Nursing Homes, Hospitals and Clinics]

Turning now to some of the additional characteristics of the supply of nursing labour, Table 1.4 shows some of the supply-related demographic characteristics of the NHS nursing workforce based on the NHS Hospital and Community Health Services Non-Medical Workforce Census. The main points are that the majority of qualified nurses are female, between the ages of 25 and 54 years, and are of white ethnic origin. Interestingly, note that while only around 15% of nurses are from non-white ethnic groups this is probably a higher proportion than for other occupations. For example, surveys such as the Quarterly Labour Force Survey (QLFS) suggest that the proportion of workers in occupations other than nursing who are of non-white ethnic origin is approximately two thirds the proportion of nurses who are non-white (see Table 5.1 in Chapter 5 for information from the QLFS).

	Qualified [1,2]	Unqualified [1,2]	Learners [1,2]	Total [1,2]
Total WTEs [3]	256,280	89,830	1,970	346,180
Sex				
% Male	11.8	15.4	9.1	12.7
% Female	88.2	84.6	90.9	87.3
Total	100.0	100.0	100.0	100.0
Age				
% <25 years	4.2	6.7	8.1	4.9
% 25-34 years	27.5	18.7	46.7	25.2
% 35-44 years	34.3	24.8	34.5	31.8
% 45-54 years	23.5	25.5	9.6	24.0
% 55-64 years	6.7	14.1	0.5	8.6
% >65 years	0.0	0.1	0.0	0.1
% Unknown	3.7	10.1	0.5	5.4
Total	100.0	100.0	100.0	100.0
Ethnic group				
% White	86.0	82.0	86.7	85.0
% Black	4.8	4.1	5.9	4.6
% Asian	1.8	1.2	1.2	1.7
% Other	2.7	2.2	2.1	2.5
% Unknown	4.7	10.5	4.1	6.2
Total	100.0	100.0	100.0	100.0

[1] Figures represent characteristics of staff in England as at 30th September 2000.
[2] Includes nurses, midwives and health visitors.
[3] Whole-time equivalents

Table 1.4. Demographic characteristics of nursing, midwifery and health visiting staff in England, 2000 [source: Department of Health. NHS Hospital and Community Health Services Non-Medical Workforce Census

A notable feature of occupations that are populated primarily by females is the significance of part-time working. This is a direct effect of traditional family role specialisations. One would therefore expect in a labour market which is predominantly (around 90%) female there to be a significant proportion of part-time workers. Table 1.5 shows the ratio of numbers of staff to whole-time equivalents. In 2000, for example, there were 1.24 qualified nursing staff employed for every whole-time equivalent post. Put another way each qualified nurse working in the NHS worked on average 0.8 of a whole-time equivalent post.

	1995	1996	1997	1998	1999	2000
Qualified [1]	1.21	1.21	1.22	1.23	1.24	1.24
Unqualified [1]	1.36	1.35	1.39	1.38	1.37	1.38
Learners [1]	1.04	1.09	1.04	1.03	1.03	1.04
Total [1]	1.25	1.25	1.26	1.27	1.27	1.27

[1] Includes nurses, midwives and health visitors.

Table 1.5. Ratio of numbers (head count) of nursing, midwifery and health visiting staff to whole-time equivalents in England, 1995-2000 [source: Department of Health. NHS Hospital and Community Health Services Non-Medical Workforce Census]

1.4. Nurses' pay determination

Having identified the main components of demand and supply in the nursing labour market we now examine how the prevailing wage rate is determined through the interaction of employers, who demand nursing labour, individual nurses, who supply their time and effort in return for payment at the market wage, and the Pay Review Body. In the following chapters we examine the nature and magnitude of nurses' earnings given this process of pay determination.

As suggested above the determination of nurses' pay is far from straightforward, illustrating clearly the difficulties in setting wages in a market with externally imposed resource constraints and a monopsonistic employer. Due to the imperfect nature of the market competitive forces are unable to determine the wage rate and due to measurement difficulties nurses' pay cannot be linked to marginal productivity. This situation is further complicated by the fact that nurses generally work in the public sector and their pay is determined in the much larger context of general public expenditure. These problems are reflected throughout the history of nurses' pay determination. Before an analysis of nurses' earnings may proceed it is important to understand the method by which their pay levels are determined under these conditions. Because of the way that the system has evolved over time it is useful to place the discussion in a historical context. The key events in the history of nurses' pay determination are presented in Table 1.6. [4]

1.4.1. The Nurses and Midwives Whitley Council, 1948–1983

Except for the period 1995-1996 when an element of local pay bargaining was introduced NHS nurses' pay has always been determined at the national level. From 1948 to 1983 nurses' pay was determined primarily by the Nurses and Midwives Whitley Council. Nurses and their representatives (the Staff Side – the unions and professional organisations) and managers and employers and their representatives (the Management Side – the Departments of Health and NHS management) met regularly to negotiate pay structures and pay increases within the profession as well as to agree other terms and conditions of employment.

1948	The Nurses and Midwives Whitley Council is established
1967-1968	Independent pay review for nurses by the National Board for Prices and Incomes
1973-1974	Independent pay review for nurses by the Committee of Inquiry into the Pay and Related Conditions of Service of Nurses and Midwives (Chairman: Lord Halsbury)
1979-1980	Independent pay review for nurses by the Standing Commission on Pay Comparability (Chairman: H. A. Clegg)
1983	The Pay Review Body for Nurses, Midwives and Health Visitors is established
1987-1988	Clinical Grading Review
1995-1996	Local pay determination is introduced

Table 1.6. Events in nurses' pay determination, 1948-2001

[4] See Gray (1989) for an in-depth review of this subject.

There were recurring difficulties with this system, with both sides rarely reaching agreement on pay increases. This is understandable since the Whitley Council framework essentially relies on different parties with fundamentally opposite aims and objectives reaching an agreement on pay increases. The inability of the Whitley Council to reach agreement often led to arbitration or to the involvement of the industrial court. The outcomes were twofold: either low pay increases were awarded; or, the length of time before an agreement was reached was so long that no changes to nurses' pay were made in a given year. Because of the cumbersome nature of this system there was continuous recourse to special reviews conducted internally by the Whitley Council to relieve the tension and to allow improvements in nurses' pay levels. There were four such 'special reviews' in the 1960s (Committee of Inquiry into the Pay and Related Conditions of Service of Nurses and Midwives, 1974) but they had the effect only of upgrading previous pay levels so that they became neutral in real pay terms, rather than awarding actual increases in real pay. These catch up exercises, which sought to redress the imbalance in the system, were indicative of the difficulties with the Whitley Council structure.

The first independent review of NHS nurses' pay conducted outside the Whitley Council system was undertaken in 1967-1968 by the National Board for Prices and Incomes. The specific remit of the Board was to adjudicate the Staff Side's request for a substantial increase in pay, along with additional requests for a reduction in hours worked and the extension of special duty payments. Following an extensive survey of conditions in more than 500 hospitals the Board recommended pay rises of 9-14% across the profession. Following this one-off pay hike in 1968 a second independent body was asked in 1973 to resolve an impasse between the Staff Side and Management Side over pay. The Committee of Inquiry into the Pay and Related Conditions of Service of Nurses and Midwives under the chairmanship of Lord Halsbury was appointed to examine the pay structure and levels of remuneration and related conditions of nurses and midwives. It had two broad aims – to establish a general pay level that was appropriate for nurses in relation to other groups, and to rationalise the pay structure. Following evidence submitted from the Staff Side and Management Side and on their own survey data the Committee of Inquiry recommended a substantial increase in nurses' pay of, on average, 30%. All parties seemed satisfied with the outcome, but cutbacks in government expenditure in the mid-1970s plus the high rate of inflation meant that this pay increase was eroded over a very short period of time. Therefore the Staff Side submitted a claim to the government in 1978 for special treatment and the Standing Commission on Pay Comparability under the chairmanship of H.A. Clegg was appointed 1979 to make recommendations for an uplift in nurses' pay. The Standing Commission undertook an extensive analysis of nursing work comparing the responsibilities, training requirements and levels of remuneration of nurses with those of workers in other occupations. The two main outcomes of the review were an increase in real pay by an average of 18% across the profession and a reduction in the length of the normal working week for qualified nurses from 40 to the current 37½ hours. Once again, after a sizeable one-off increase pay rises returned to the lower levels that were routine under the Whitley Council system. Following another standoff in pay negotiations in the early 1980s which resulted in nurses taking industrial action the government agreed to establish in 1983 the independent Pay Review Body for Nurses, Midwives and Health Visitors.

1.4.2. The Pay Review Body

Under the Whitley Council regime there was considerable disagreement over the magnitude of nurses' pay rises, with the result that pay levels often fell or at best stayed level in real terms year on year. Various catching up exercises were undertaken in the form of independent reviews. Each of these reviews was a one-off exercise, each had an immediate positive effect on pay levels, and the effect of each was then eroded in succeeding years. This 'stop-go' situation finally ended in 1983 with the establishment of the Pay Review Body, which is still in use today.

The remit of the Pay Review Body is to advise the Prime Minister on the pay of qualified and unqualified nurses, midwives and health visitors, including nurse trainees working in the NHS (the latter are usually referred to as learners). As noted in its first report: "We are an independent body and our task is to recommend appropriate remuneration for nursing staff in the light of all relevant factors. On the one hand we have an obligation to consider what is fair to the nursing staff themselves. On the other hand we must also have regard to the interests of the taxpayer and to the general economic situation. But the community also has an interest as users of health care in having an efficient National Health Service manned with appropriately trained, experienced and motivated staff. Our primary aim for nursing staff, midwives and health visitors is, therefore, to establish a stable system of pay determination which will ensure fair levels of remuneration for the nursing profession and safeguard the interests of the community both as taxpayers and as users of health care." (Pay Review Body for Nurses, Midwives and Health Visitors, 1984).

Compared to the previous one-off independent review committees the Pay Review Body also declared that it had what it called a "major advantage": "we are a standing body and will be keeping the pay of nursing staff under review continuously from year to year. We shall therefore be able progressively to take account of a wide range of factors including job content and organisation, pay developments elsewhere, and changing economic circumstances."

Nevertheless the operation of the Pay Review Body is similar to that of the previous one-off independent review committees. It reviews evidence from nurses and their representatives (the Staff Side), managers and employers and their representatives (the Management Side) and from the wider economy in terms of labour market conditions. The review process itself follows an annual cycle. The Pay Review Body examines evidence available (or submitted) to it up to December of any given year, and then forwards its recommendations on nurses' pay and publishes its report the following year with an expected full implementation date of April 1st. As stated above the Pay Review Body is independent. It makes its own independent recommendations on nurses' pay based on the evidence. The recommendations are not directly cash limited and unlike the previous system the process cannot be disrupted by delaying tactics. It is also important to note that the government is obliged to accept its recommendations. As stated in the first Pay Review Body report: "[t]he Prime Minister has made it clear in the Chairman's letter of appointment that 'successive Governments have agreed to accept Review Body recommendations'" although as is then noted, this is unless "'there are clear and compelling reasons for not doing so.'" In fact the government has always implemented the recommendations made by the Pay Review Body though in some years the recommended pay increases have been delayed or staged. This can be seen more clearly with reference to Table 1.7 which shows the magnitude of the Pay Review Body's recommendations (in terms of the percentage increase in wages based on changes to the salary scales) and the extent to which they have been implemented. In summary the government has staged or delayed full implementation in six of the seventeen years that the Pay Review Body has been reporting (staging also occurred in two further years due to inadequacies with local pay bargaining – see below). In these six years as noted by Buchan (1995) the government has trodden a fine definitional line. It can claim to have implemented the Pay Review Body's recommendations since it has done so by the end of the year in question. However, to its advantage a delayed or staged implementation reduces the paybill by tens of millions of pounds for that year. Nurses have realised salary increases each year under the Pay Review Body system. [5] In some years these have been substantial, though the effects have been limited somewhat by government intervention delaying or staging full implementation. With this broad picture in mind we now consider two further developments in nurses' pay determination, namely the clinical regrading exercise that was conducted in 1987-1988 and local pay bargaining which was introduced in the period 1995-1996.

Year	Pay Review Body recommendation (% increase) [1]	Extent of implementation
1984	7.0	Fully implemented
1985	8.6	Staged by government
1986	7.8	Delayed by government
1987	9.5	Fully implemented
1988	15.3	Fully implemented
1989	6.8	Fully implemented
1990	9.6	Staged by government
1991	9.7	Staged by government
1992	5.8	Fully implemented
1993	1.5	Fully implemented
1994	3.0	Fully implemented
1995	3.0	Staged due to local pay bargaining
1996	2.8	Staged due to local pay bargaining
1997	3.3	Staged by government
1998	3.8	Staged by government
1999	4.7	Fully implemented
2000	3.4	Fully implemented

[1] Based on the mean change in salary scales

Table 1.7. Pay Review Body recommendations concerning nurses' pay, 1984-2000

[5] In the next chapter we examine trends in nurses' earnings over time, including the impact of the Pay Review Body.

The pay and grading structure for NHS nurses remained basically the same from the formation of the NHS in 1948 until 1988. In that year a new career structure within the nursing profession was implemented, with the new nursing grades being priced (i.e. assigned a salary level) by the Pay Review Body. The aim of the exercise was to provide clear job descriptions for all nursing posts within the NHS so that roles and tasks could be standardised. Pay was then linked by the Pay Review Body to the tasks performed. A new alphabet structure was introduced which is still used today. Unqualified nursing staff were placed on grades A and B and enrolled nurses were placed mainly on grades C and D. Registered staff nurses were placed mainly on grades D and E, and ward sisters were placed on grades G and H. Each grade was awarded its own six-point pay scale with a 3-4% differential between each point. There was minimal overlap between scales for each grade.

With the pricing of the new structure by the Pay Review Body and the input of extra funds by the government to meet the cost of the regrading the nursing profession on average experienced a substantial one–off increase in pay of 15.3%. However, while this was the average effect there were individual gainers and losers which created a number of problems. At the local level nurses who previously had carried out many of the same tasks were now assigned to different grades by their employers. Given the salary scales attached to each grade by the Pay Review Body this meant that nurses with essentially the same jobs were awarded different pay increases. This caused much discontent within the profession. The result was that many nurses appealed against their regrading outcome, though the appeals procedure was not sufficient to cope with the volume. The upshot was that over three years later more than 30,000 appeals were still waiting to be heard (Pay Review Body for Nurses, Midwives and Health Visitors, 1992).

More recently a major change in nurses' pay determination was the introduction of local pay bargaining during the period 1995-1996. Since 1987 the Management Side expressed to the Pay Review Body a wish to introduce more flexible pay arrangements in order to enable the NHS to adjust to differences in local labour market conditions (i.e. the geographically uneven distribution of qualified nursing shortages). This was couched in terms of being able to offer higher salaries to nurses in areas where there was a nursing shortage. A number of pilot schemes were initiated over the period 1987 to 1994 and the Pay Review Body signalled a change of approach in 1994. In its 1995 report the Pay Review Body then directed employers and nursing staff to engage in local pay bargaining. The upshot was that in its 1995 report the Pay Review Body recommended a two-tiered approach to awarding pay increases: a 1% national increase in salary scales across the profession supplemented by additional pay rises to be determined by local negotiations on pay. As a guide the Pay Review Body declared that the outcome of local pay negotiations should lead to an additional pay rise of at least ½-2% over and above the change in national rates. In practice few local pay settlements were reached in the 1995 round of pay negotiations. This was partly a reflection of the time taken by local employers to draw up business plans for local pay determination. Where local bargaining was undertaken nearly 90% of settlements were for an additional 2% – the highest amount originally recommended by the Pay Review Body. As a consequence one year later in 1996 the Pay Review Body uprated the 1994 national pay salaries by a total of 3% reflecting the 1% it initially recommended plus an additional 2%. In summary the local pay bargaining system did not operate properly in its first year, and this had the added effect that the 3% increase in pay for 1995 was staged, effectively due to the inadequacies of local pay bargaining.

In the second year of local pay bargaining (1996) the Pay Review Body recommended a 2% increase in national salary scales plus a further increase to be negotiated locally. The magnitude of the additional increase was not indicated, as in the previous year, but it was noted that the government was to provide 3.9% of additional funding for hospital and community health services. The response by local employers was disappointing from the point of view of the Pay Review Body. As it noted in its 1997 report "the process operating in 1996 has generally left the local element of pay for nursing staff as a residual to be determined after all other demands on available funds have been satisfied. The outcome for individual nursing staff has been uncertain and often very slow to materialise, and it is not how we envisaged matters proceeding when we made our recommendations." The problem was that local employers generally took the view that pay increases should equal the rate of inflation and therefore be neutral in real terms. There was little negotiation between employers and staff. The outcome was that once again the Pay Review Body intervened and uprated salary scales nationally with the effect that the 2.8% total increase was staged as a consequence of local pay bargaining for the second year running.

Following this lack of success in 1997 the Pay Review Body changed tactics and reverted back to the old system, recommending a 3.3% national increase. To all intents and purposes the practice of local pay bargaining was discontinued.

Local pay bargaining was unsuccessful for two reasons: the adjustment time required by local employers to construct the machinery within which the process could operate at the local level: and, an apparent unwillingness of local employers to negotiate over pay. From the outset the Staff Side in national negotiations was vehemently opposed to the whole idea of local pay bargaining. It argued that local pay bargaining was unfair and that it undermined the role of the Pay Review Body as the guardian of fair pay in nursing. The Staff Side believed that local pay bargaining would in fact result in lower pay for the majority of nurses. Rather than eroding the monopsony power of employers at a national level it would instead increase monopsony power at the local level. Put another way rather than have a single monopsony employer held in check by a Pay Review Body there would instead be many local monopsonistic employers each able to act freely without interference by an independent review body. It was also argued rightly that the transaction costs would be high because employers would need to employ additional staff to take part in negotiations and there would be additional costs to cover nursing staff who took part in the local negotiations. These misgivings do seem to be borne out by the outcome of local pay bargaining in its two years of operation.

In summary, the history of nurses' pay determination in Great Britain can be divided into two periods. The 'Whitley Council period' from 1948 to 1983 was one of marked fluctuations in nurses' pay with internal special reviews and one-off external reviews initiated as catch-up exercises. By contrast the 'Pay Review Body period' from 1983 to the present has been characterised by less variability and comparative improvements in nurses' pay. There has been some fluctuation throughout this period, most notably from the clinical regrading exercise.

1.4.3. Nurses' pay determination under the Pay Review Body System

We have described the historical context of nurses' pay determination in Great Britain. We now describe and explain how nurses' pay is established under the Pay Review Body system. To recap, the remit of the Pay Review Body is to make recommendations on the pay levels of qualified and unqualified nurses, midwives and health visitors, including learners, working in the NHS, which the government is obliged to accept.

The Pay Review Body has made it clear that the issues it takes into account in its deliberations are as follows (Pay Review Body for Nurses, Midwives and Health Visitors, 1984, Buchan, 1995):

1. Affordability of potential pay rises and the likely impact of these on the number and types of nurses employers will be able to employ with fixed budgets;
2. Recruitment and retention in terms of vacancy rates, turnover rates and uptake of nurse training programmes;
3. Fairness and comparability of earnings with those of other workers within the NHS and elsewhere;
4. Morale and motivation; and,
5. Productivity and workload.

The Pay Review Body reviews evidence on these issues submitted from three main sources:

1. Nurses and their representatives (the Staff Side);
2. Managers and employers and their representatives (the Management Side); and,
3. From the wider economy, in terms of labour market conditions.

Typically the Staff Side in their evidence emphasise the need for fair pay for nurses, recognising nurses' training and qualifications and their roles and responsibilities in the provision of high quality health care within the NHS regardless of the costs involved (see, for example, Pay Review Body for Nurses, Midwives and Health Visitors, 2001). In recent years they have also focused on the existence of nursing shortages, coupled with what they term the 'casualisation' of the nursing workforce (the increased use of bank and agency nurses) and an over-reliance on overseas recruits to the profession. They also point out frequently that nurses' pay lags behind that of comparators and that these differences increase over the working life.

In their 2001 submission to the Pay Review Body the Management Side pointed out that recruitment and retention is healthy in the nursing profession and that improved working conditions have been achieved over time (Pay Review Body for Nurses, Midwives and Health Visitors, 2001). They did not submit evidence on pay comparability and warned against placing too much emphasis on comparisons with other employee groups because there was a risk of "cherry picking" comparators. In 2001 as in previous years the Management Side emphasised the need for pay levels that allow them to employ an adequate number of nursing staff within their limited budgets. They argued that "[t]he envisaged increase in the [nursing] workforce would ease the burden on existing staff, although funding for extra staff was reliant on pay increases for existing NHS staff being realistic and affordable" (Pay Review Body for Nurses, Midwives and Health Visitors, 2001). They also submitted that: "[t]he level of pay award would also determine how much money was available for NHS pay modernisation. Pay increases should be affordable so that modernisation could take place, as change could not be carried out without significant cost." Further, they pointed out that the government had set challenging targets in terms of improved provision of health care and that these would require a significant increase in NHS staff. They stated: "[a]s an extra five thousand nurses would add about 1.5 per cent to the nurses' paybill, increases to the workforce could only be achieved if pay awards were appropriate, realistic and affordable (Pay Review Body for Nurses, Midwives and Health Visitors, 2001). Clearly the issue of affordability is of paramount importance to the Management Side. [6]

From the wider economy the Pay Review Body reviews evidence on the rate of inflation, average earnings in the economy, and pay settlements in other occupations. As noted in the 2001 report: "We recognise the importance of the Government's inflation target. Nurses cannot expect to be considered immune to the pressures felt elsewhere in the economy. However, our objective in ensuring fair pay for nurses must involve a consideration of their place in the wider economy, and our recommendations cannot be divorced from general pay and prices movements" (Pay Review Body for Nurses, Midwives and Health Visitors, 2001).

Given this description of the mechanisms operating in the market it is now possible to examine the potential effects of the Pay Review Body system on nurses' pay. These are presented graphically in Figure 1.3. A target amount of health care may be defined at which the health care needs of the population are met. This represents the ideal amount of health care that would be provided with no resource constraints. Q in Figure 1.3 is the number of nurses required to meet this ideal. B1 represents the budget constraint imposed by the government. This is an isoexpenditure curve delineating the trade-off between wages and the number of qualified nurses employed given the size of the NHS budget and the proportion allocated to nurses' pay. The supply curve is S1.

[6] Of course in response the Staff Side constantly argues that the Pay Review Body is not bound by a budget constraint and that it should make recommendations that it believes are appropriate under the circumstances regardless of the costs involved, leaving the Government to deal with issues of affordability.

Given the submitted evidence of the kind discussed above the Pay Review Body every year makes recommendations concerning nurses' pay. In very general terms the outcome of the process will result in wages falling somewhere between a relatively the high wage level preferred by nurses and their representatives and the lower wage level preferred by employers and their representatives. More specifically in terms of Figure 1.3 where the Pay Review Body sets the wage level at the point of intersection of B1 and the supply curve S1 wages are set at W* and the number of nurses employed is N*. [7] Note that this combination of pay and employment is 'affordable' since it is not to the right of the budget line B1. The wage outcome (W*) is unlikely to be desirable to the Staff Side who as noted above argue for fair and reasonable pay for nurses that in their opinion should not be bound by a budget constraint. They argue that the Pay Review Body should make recommendations that it believes are appropriate under the circumstances regardless of the costs involved. The upshot is that if nurses and their representatives believe that the level of pay and the workload for nurses indicated by W*N* are unfair – for example if they believe the MVP curve lies to the right of B1 (e.g. MVP1 in Figure 1.3) so that the unconstrained equilibrium wage would be greater than W* (e.g. W1) – then W* will be unacceptable to the Staff Side and they will attempt to negotiate higher wages even though this would mean moving to the right of the budget line.

Note that even at the constrained equilibrium wage W* there is a target shortage relative to the needs-based optimum given by Q-N*. If the actual wage is set above W* the result is an excess supply of nursing labour. In this case the budget constraint then becomes the limiting

[7] This is not a perfectly competitive equilibrium because it is achieved by the intersection of the supply curve and the isoexpenditure curve rather than the demand curve. For this reason we shall call the point of intersection the 'constrained equilibrium'.

factor in the market – government expenditure levels effectively determine the number of nurses employed.

Another possibility is that the market wage is set by the Pay Review Body below W*. In this situation the supply curve dictates the resulting level of employment. Under monopsony conditions where the market power of the employers prevails the employer faces an upward-sloping factor supply curve (S1). The marginal expenditure on labour by employers (ME) exceeds the average cost so that the ME curve lies above S1 at all levels of employment. The monopsonistic equilibrium occurs at W2N1. This is the preference of the monopsonistic employers. Compared to the constrained equilibrium defined above the market wage will be lower and there will be fewer workers employed. Interestingly, note that in a monopsonistic setting there will be further shortages in addition to the 'needs-based' shortages described above. At monopsony equilibrium wage W2 the number of nurses employed is N1 and there will be an 'equilibrium' shortage of N2-N1. In addition there will also be a needs-based shortage of magnitude Q-N2. Note that the equilibrium shortage does not represent an excess demand because the monopsonistic employer will not be willing to pay higher wages, though they will be willing to employ more workers at the going wage rate. The reported shortage will not exert an upward pressure on wages. The important point is that under monopsony conditions a reported shortage of workers may be perfectly consistent with the labour market being in equilibrium. [8]

[8] See Yett (1973) for further discussion of the nature of nursing shortages

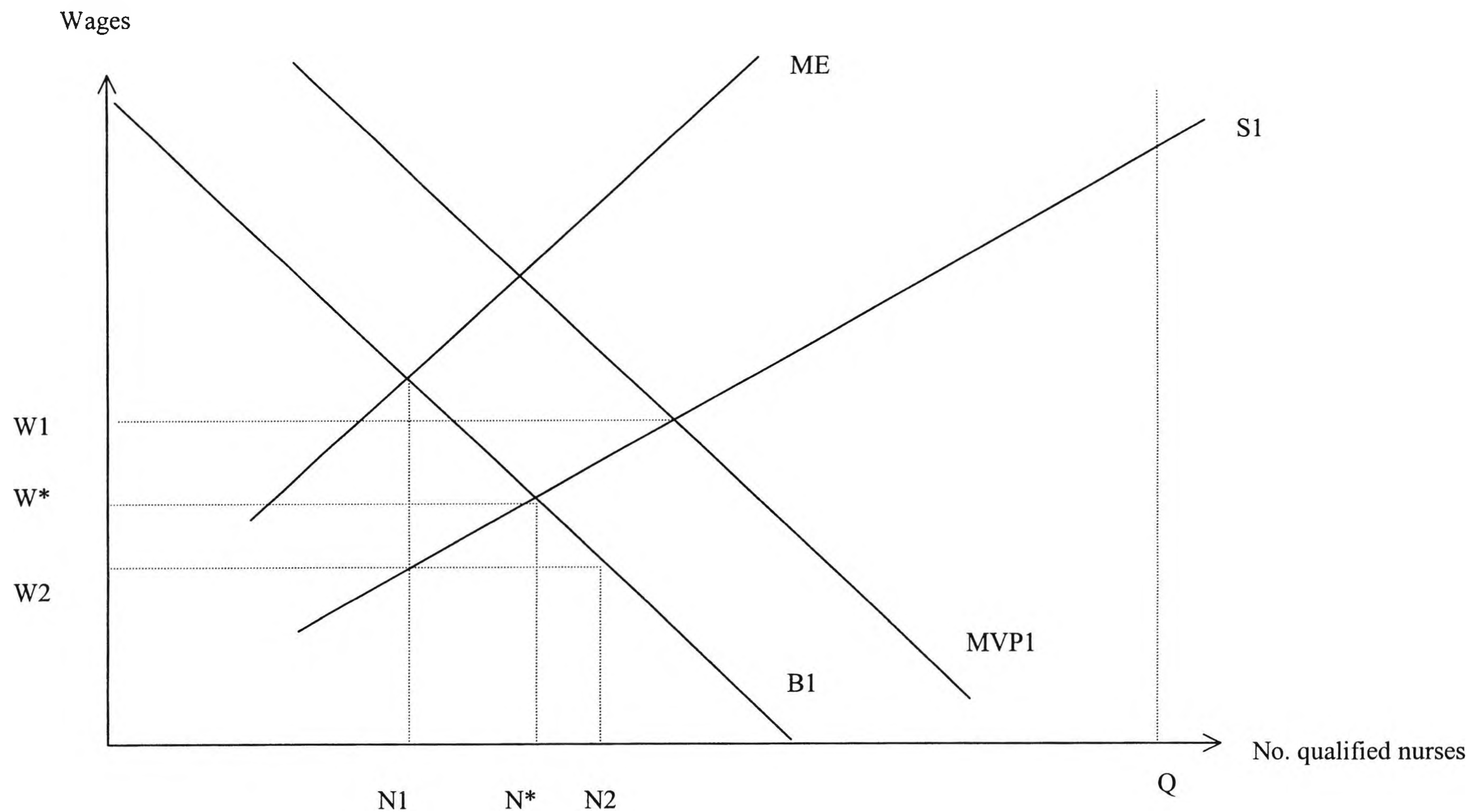


Figure 1.3. The interaction of demand and supply in the nursing labour market

The upshot from the above discussion is that the likely outcome of nursing pay negotiations is a wage level set at or above W2 in Figure 1.3. On the one hand the Staff Side argue continually for higher wages reflecting fair pay for nurses and on the other hand the Management Side ask that the pay levels recommended by the Pay Review Body be affordable given current resource constraints and the number of nurses required to provide minimum standards of health care: the Management Side's preference is for wage levels tending towards W2; the Staff Side's preference is for a higher wage, for example of the order of magnitude W1. The actual level achieved will depend on the strength of the evidence submitted to the Pay Review Body by each side and the issues the Pay Review Body deems important when making its recommendations.

Given this background it now remains to be seen which factors influence the Pay Review Body most importantly in making its recommendations. Morris (1998) notes that the impact of the evidence submitted to the Pay Review Body by nurses and their representatives is likely to be greater in the following circumstances (note that the reverse is required for the bargaining strength of employers and their representatives to be greater):

1. When there is little scope for substituting other factors of production for nursing labour, so that nurses are irreplaceable;
2. When the provision of health care is unaffected by the wage rate payable to nurses, so that increasing the wage rate will have little impact on the provision of health care;
3. When labour costs are a small proportion of the total costs of providing health care, so that increasing the wage rate will have little impact on the total costs of providing health care;

4. When the collective body of nurses is strong, so that it is unified and has significant resources at its disposable;
5. When health care providers earn substantial profits, which may be devoted to higher wage rates;
6. When nurses are prepared to offer productivity deals in compensation, for example, by working longer hours in return for increased wages.

As noted above the Management Side in their submissions to the Pay Review Body frequently emphasise that increasing the wage rate will have a substantial impact on the provision of health care via the number of nurses they can afford to employ (implying that point 2 in the list above does not hold) and that nursing labour costs are a large proportion of the total costs of providing health care, so that increasing the wage rate will have a major impact on the total costs of providing health care (implying that point 3 does not hold). They argue that affordability is a key issue – in terms of Figure 1.3 the argument is that the combination of wages and employment prevailing in the nursing labour market should fall on or to the left of the isoexpenditure curve B1. The Staff Side on the other hand constantly argues that the Pay Review Body is not bound by a budget constraint and that it should make recommendations that it believes are appropriate under the circumstances regardless of the costs involved. The Pay Review Body acts on the principle that it's role is to award levels of pay that are fair, both to the nursing profession themselves *and* to the general community as a whole. The implication is that large pay rises might be fair for nurses, but may be less so for the community if it means that less nurses will be employed to provide health care.

Evidence is available concerning the relationship between nurses' earnings and NHS expenditure. In the year 2000 total NHS expenditure was £57 billion (Office for Health

Economics, 2001). The total nursing paybill was £10 billion (Pay Review Body for Nurses, Midwives and Health Visitors, 2001). Figure 1.4 shows the relationship between year on year between changes in nurses' earnings and changes in NHS expenditure.

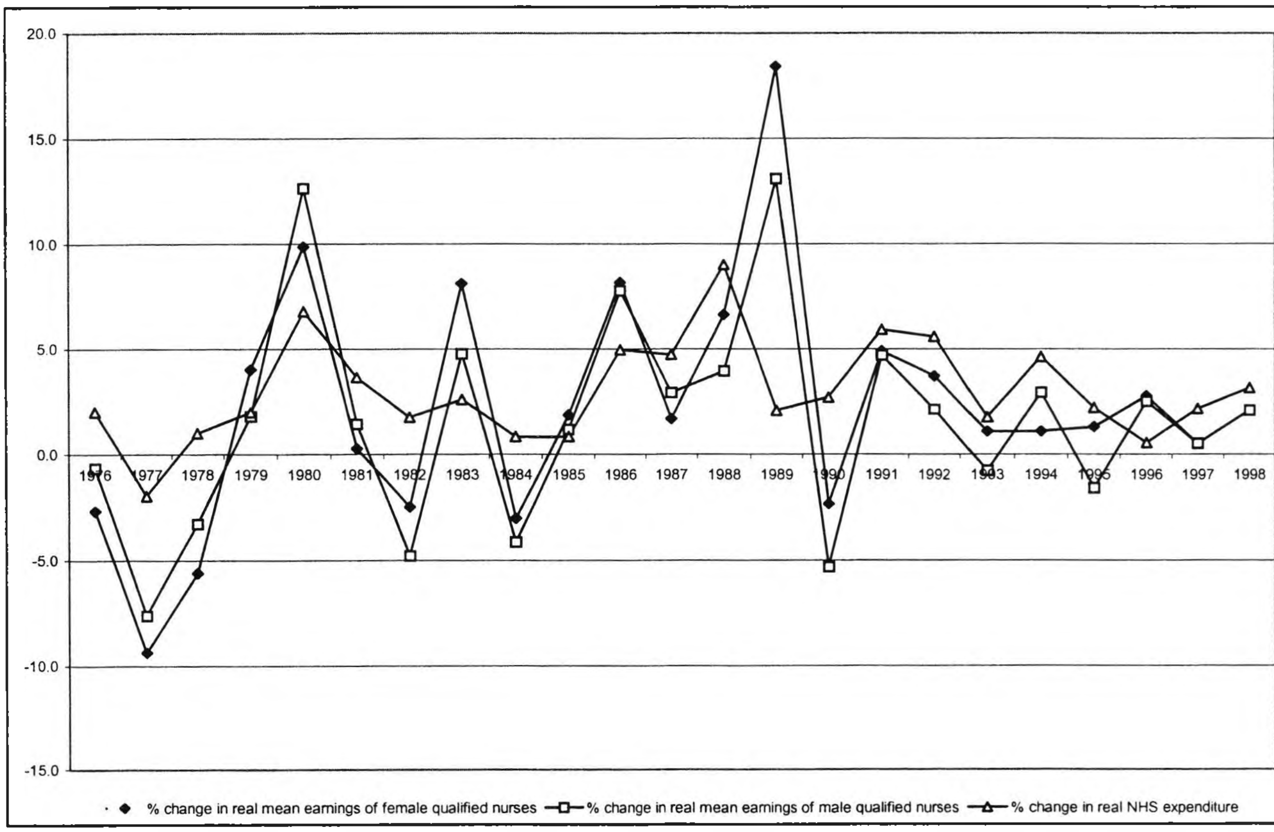


Figure 1.4. % Change in NHS expenditure and % change in nurses' earnings, 1975-1998.

There is generally a close positive relationship and, except for 1989 – the year in which the clinical regrading review was conducted and nurses' on average received a substantial increase in pay – changes in earnings are closely linked to changes in NHS expenditure. This is indicative of the emphasis the Pay Review Body gives to the issue of affordability in setting nurses' pay (contrary to the preferences of the Staff Side). As we noted above the magnitude of NHS expenditure will ultimately determine the size of the budget given to managers and employers. In pay negotiations the Management Side continually argue for small increases in nurses pay and emphasise the need for pay levels that allow them to employ an adequate number of nursing staff within their limited budgets. They stress repeatedly the importance of affordability and the effect that substantial increases in nurses' pay will have on their ability to employ the appropriate level of nursing staff. As the Pay Review Body noted recently: "[t]he [Management Side] asked us to make a straightforward, realistic *and affordable* generic increase to basic pay rates" (Pay Review Body for Nurses, Midwives and Health Visitors, 2001 [emphasis added]). The relationship shown in Figure 1.4 suggests the importance the Pay Review Body gives to concerns regarding affordability raised by the Management Side because changes in nurses' pay are closely linked to changes in NHS expenditure (budget).

Further evidence on the outcome of the Pay Review Body's deliberations is provided in Figure 1.5, which shows the change in annual salaries, change in whole-time equivalent numbers and vacancy rates for qualified staff over the period 1984 to 1997. [9] One important trend is discernible: across the entire period there is a direct mapping (a positive relationship) between fluctuations in pay and whole-time equivalent numbers of qualified staff. Put simply, the percentage change in real salaries is reflected by similar changes in the numbers employed. Across the 1980s growth rates for pay are positive and they are also positive for the number of whole-time equivalent nurses employed. In the 1990s there is generally negative growth in real pay, and this is also mirrored by negative growth in numbers. Because larger amounts of labour are being employed at higher wages and vice versa this is suggestive of supply being the dominant factor in the labour market and the supply curve driving the relationship between wages and employment. The implication is that market wages are being set by the Pay Review Body more in line with lower wage levels preferred by employers rather than the higher levels preferred by the Staff Side.

[9] Note that the trends in whole-time equivalent numbers employed shown in Figures 1.2 and 1.4 are directly comparable.

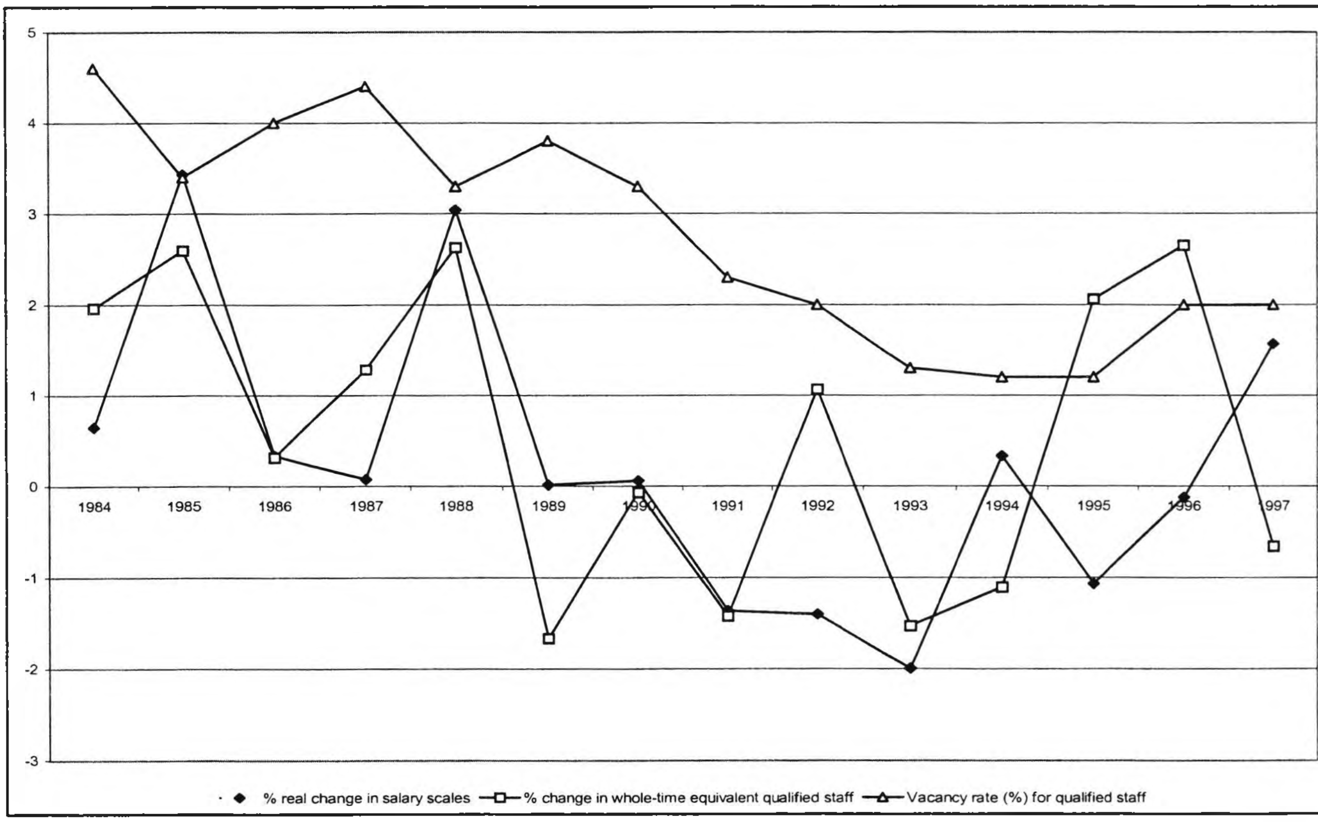


Figure 1.5. Growth in annual salary, growth in whole-time equivalents and vacancy rates for qualified nurses, midwives and health visitors in Great Britain, 1984-1997 [source: Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine].

The evidence presented above suggests that the issue of affordability propounded by the Management Side is deemed to be important by the Pay Review Body in its deliberations. More recently the picture appears to be changing. In its 2001 report the Pay Review Body stated: "We have noted the evidence on recruitment and retention. This gives some very mixed messages and, in our view, it would be very unwise to claim that problems are even close to resolution. Evidence on motivation and morale shows some improvement in a range of measures, but suggests there is still some way to go.... We note also the arguments from the NHS employers that our award this year should have as an objective the freeing up of funding for extra staff. We understand and sympathise with the arguments behind this viewpoint. But we consider that a full solution to the human resource problem affecting the

NHS still lies a little way off. In the meantime it makes sense to us to ensure maximum retention of existing staff, as well as maintaining attractive opportunities for new entrants and returners" (Pay Review Body for Nurses, Midwives and Health Visitors, 2001). The outcome was a recommended 3.7% increase in pay rates over the previous year compared to 3.4% in the previous year (see Table 1.7).

Vacancy data provide a measure (albeit an imperfect one) of the persistence of shortages. As part of the evidence on which it bases its recommendations concerning recruitment and retention the Pay Review Body has considered the results of surveys of nursing vacancies conducted annually by the Office of Manpower Economics of all health care providers in Great Britain that are known to employ nursing staff. Results of these surveys, which were conducted between 1984 and 1997, are also presented in Figure 1.5. The vacancy rates presented are calculated as the number of posts unfilled for three months or more as a percentage of the total number of qualified staff in post. While these percentages do appear to be small (1% to 4%), when one considers the size of the qualified nursing workforce, the numbers of unfilled nursing posts indicated by these vacancy rates becomes significant in absolute terms. The general trend across the period is that vacancies are falling, with an upturn towards the end of the period. It should be borne in mind that vacancies are an unreliable measure of the nursing shortages due to problems of definition and the fact that if the vacancy rate is low and fairly stable over time (as is shown generally to be the case here) then in large organisations such as NHS Trusts it is predictable and can be incorporated into planning requirements. More recent estimates place the actual nursing shortage at closer to 15,000 (Hancock, 1999). The implication is that the Pay Review Body has in the past set wages more in line with the level preferred by the Management Side (e.g. at or below W* in Figure 1.3) rather than at the higher levels preferred by the Staff Side (e.g. W1). In the longer term persistent shortages might cause the Pay Review Body to re-evaluate the importance of recruitment and retention relative to affordability as the central principle in setting pay levels. Evidence of such a reassessment is given in the 2001 report discussed above.

1.5. Conclusions

In this chapter we have examined the mechanisms by which nurses' pay is determined in Great Britain. The nursing labour market has a complex market structure and we have demonstrated the unusual and complicated nature of the market forces in operation. In particular it is worth emphasising the crucial role of the government in setting the NHS budget and, as a consequence, defining the expenditure limits within which wage and employment decisions are constrained. Also important on the demand-side of the market is the Pay Review Body that determines levels of pay, but whose recommendations are influenced by the monopsony power exerted by NHS employers. On the supply side we find that the supply decisions of individuals in terms of joining and leaving the profession are paramount in determining the state of the labour market. The labour supply decisions of nurses and potential new entrants into the profession at the given wage are dominant features of this labour market. As discussed in section 1.3 the labour supply decisions are separable into two related choices: a participation decision; and, an occupation selection decision. We utilise this framework in subsequent chapters to analyse the determinants of nurses' pay.

In Section 1.4 we discussed the process by which nurses' pay is determined. The upshot of the discussion is that the pay levels are recommended by the Pay Review Body based on the strength of the evidence submitted from the Staff Side, the Management Side and from the wider economy. While the Pay Review Body takes a number of issues into account in its deliberations (affordability of potential pay rises, recruitment and retention, fairness and comparability, morale and motivation, and productivity and workload) the evidence suggests that in the past the issue of affordability stressed frequently by the monoponistic employers is given much prominence, though more recently it is recognised that recruitment and retention are of prime importance. The outcome is that at least up until recently the market wage rate has been set by the Pay Review Body more in line with the preferences of the Management Side, as opposed to the higher wage levels preferred by the Staff Side.

While it is noted that the nursing labour market is exceedingly complicated the exposition in this chapter is itself an oversimplification of the true market structure. Of particular importance here is the relevance of local labour market demand and supply conditions. On the demand side we noted that employers base their employment decisions on the interaction of wages and a hard budget constraint limiting the maximum number of nurses that may be feasibly employed at any give wage. While currently wage rates are set at the national level the size of the budget constraint and the number of nurses required at each grade is determined locally. On the supply side we noted that the decisions of individuals in terms of joining and leaving the profession are paramount in determining the state of the labour market. Clearly the total supply of labour (i.e. the pool of workers who are willing and able to work in a particular time period) will differ across geographical areas due to regional differences in the size and composition of the population. This will cause local differences in the supply of nursing labour (Buchan, 1995, Morris, 1998). The upshot is that the complicated market structure discussed at length in this chapter will be affected by prevailing supply and demand conditions in local labour markets, and these are likely to permeate and modify the issues raised in this chapter.

Nonetheless what the above discussion highlights is that given the nature of the supply and demand factors that persist in the market, and also given process of nurses' pay determination wages appear critical in determining the prevailing conditions. In the following chapters we proceed to analyse the magnitude and determinants of the earnings that nurses receive as the outcome of the pay-setting process described in this chapter and explain the relative earnings of nurses and workers in other occupations. We continue our analysis in Chapter 2 by reviewing how nurses' relative earnings have changed over time. First we illustrate directly the impact of the Nurses and Midwives Whitley Council, the various independent reviews and the Pay Review Body on nurses' earnings over time, as discussed in this chapter. In addition we also present a comparison of nurses' earnings with those in other occupation groups.

CHAPTER 2

A REVIEW OF NURSES' EARNINGS

2.1. Introduction

In Chapter 1 we examined the main characteristics of the labour market for qualified nurses in Great Britain and discussed how nurses' pay is determined in the NHS. With this background in mind we now conduct a review of nurses' earnings in Great Britain. We review nurses' actual earnings over time, examining how they have changed due to historical events in nurses' pay determination. We also compare nurses' earnings with earnings of workers in other occupations. We find, for example, that on average female nurses who comprise 90% of the workforce earn higher wages than other female non-manual workers.

2.2. A review of nurses' earnings

As noted by Buchan (1995) analysis of pay trends is usually an inexact science. Choice of starting dates, selection of comparator groups, definitions of 'wages', 'pay', 'salaries' and 'earnings' can vary, and the interpretation of results is often subjective. With these caveats in mind in this section we examine what has been happening to qualified nurses' earnings in Great Britain since 1975. Using data derived from the New Earnings Survey we examine trends in real weekly earnings of qualified nurses over time and show that they have generally increased year on year. Those years in which there was a marked improvement in the nurses' pay correspond with those in which nurses' pay was reviewed independently. We also examine how nurses' earnings have changed relative to those in other occupations and show how the relativities have remained reasonably constant over time.

All earnings data were computed from the New Earnings Survey using annually published occupation-specific earnings data (New Earnings Survey, Part D: analyses by occupation. ONS, selected years). The New Earnings Survey is an annual survey covering 1% of all employees in employment. All earnings data is for average gross weekly earnings of full-time workers receiving adult pay rates, excluding those whose pay was affected by absence. There are limitations in the New Earnings Survey data, particularly in terms of data on earnings of part-time workers. It is worth bearing in mind that earnings figures based on full-time workers should be treated with caution, particularly for female-dominated occupations such as nursing. As noted in Chapter 1 many nurses work on a part-time basis and their annual earnings will therefore be lower than the estimates presented here.

Nominal earnings are adjusted to constant 1975 prices. 1975 is chosen as the base year mainly because this was the first year in which comprehensive New Earnings Survey data were readily available. This was also the year in which the pay increases recommended by the 1974 report of the Committee of Inquiry into the Pay and Related Conditions of Service of Nurses and Midwives were implemented and first visible in the survey.

Adjustment to constant prices was made using the retail price index obtained from the Office of National Statistics.

At the outset it is worth noting that at the individual level nurses' earnings rise over time for two reasons. First, due to increases in nurses' salary scales. This type of increase is a result of the recommendations made every year by the Pay Review Body, as discussed in Chapter 1. Second, earnings in the nursing profession also generally rise with experience as nurses move to a higher point on the pay scale every year. This second factor might occur either as nurses are promoted to higher grades within the profession or because they move to a higher spinal point within each grade. The earnings data reviewed in this section reflect both these factors. The pay data analysed up until now have been based mainly on changes in salary scales only and therefore only include the first effect.

A second point worth bearing in mind is that the summary statistic used to measure nurses' earnings in this section is the mean. This reflects average earnings across all ages and all nursing grades and levels of experience in the profession (we do distinguish between males and females, however). This obviously reflects only at the aggregate level the earnings of the highest- and lowest-paid, and loses some of its usefulness if there is a wide dispersion in earnings. To put this into context most qualified nurses are employed on grades D to G. The newly qualified are employed at grade D and most ward sisters are usually employed at grade G. In terms of the salary scales, as of April 1st 2001 this implies a range in annual earnings of £9,975 per annum, from the bottom of the payscale for grade D (£15,445 per annum) to the top of the payscale for grade G (£25,420 per annum). There are also a range of other payments such as overtime payments, special duty payments, pay-related and non-pay-related allowances and London allowances that may widen the dispersion further.

Figures 2.1 shows the weekly earnings of male and female qualified nurses for the period 1975 to 2000. This reveals a general upward trend in earnings over this period. In 1975 following the implementation of the recommendations of the Committee of Inquiry into the Pay and Conditions of Service of Nurses and Midwives nurses' pay increased markedly from that in previous years. Real pay then fell until 1979 when it increased again following the publication of recommendations made by the Standing Commission on Pay Comparability. Earnings then changed little until 1984, which was the year in which the first report by the Pay Review Body was published. Nurses' earnings have subsequently increased in real terms each year under the influence of the Pay Review Body. In 1989 there was a marked increase in average earnings due to the implementation of the clinical regrading exercise and subsequent pricing of the new structure by the Pay Review Body. In 1990 real pay decreased slightly and returned to the slower rate of increase initiated by the Pay Review Body.

As noted in Chapter 1 around 90% of qualified nurses are female. Figure 2.1 shows that male qualified nurses receive higher earnings on average than their female counterparts. Because male and female nurses are paid according to the same pay scales the only explanation for this is that the average male nurse is on a higher nursing grade or spinal point than the average female nurse. This might arise for one of two reasons: first, because male nurses have on average more experience than female nurses; or second, because male nurses are on average in positions of greater official responsibility than female nurses. These two reasons may be related, though this will not necessarily be the case. It is worth bearing in mind that male nurses might be in positions of greater official responsibility than female nurses due to differences in individual characteristics (e.g. differences in education, ability and qualifications, differences in the amount of time spent out of nursing), but possibly also for less benign reasons such as labour market discrimination. It is difficult to draw further conclusions without detailed information on nursing workforce composition by age, sex and pay grade.

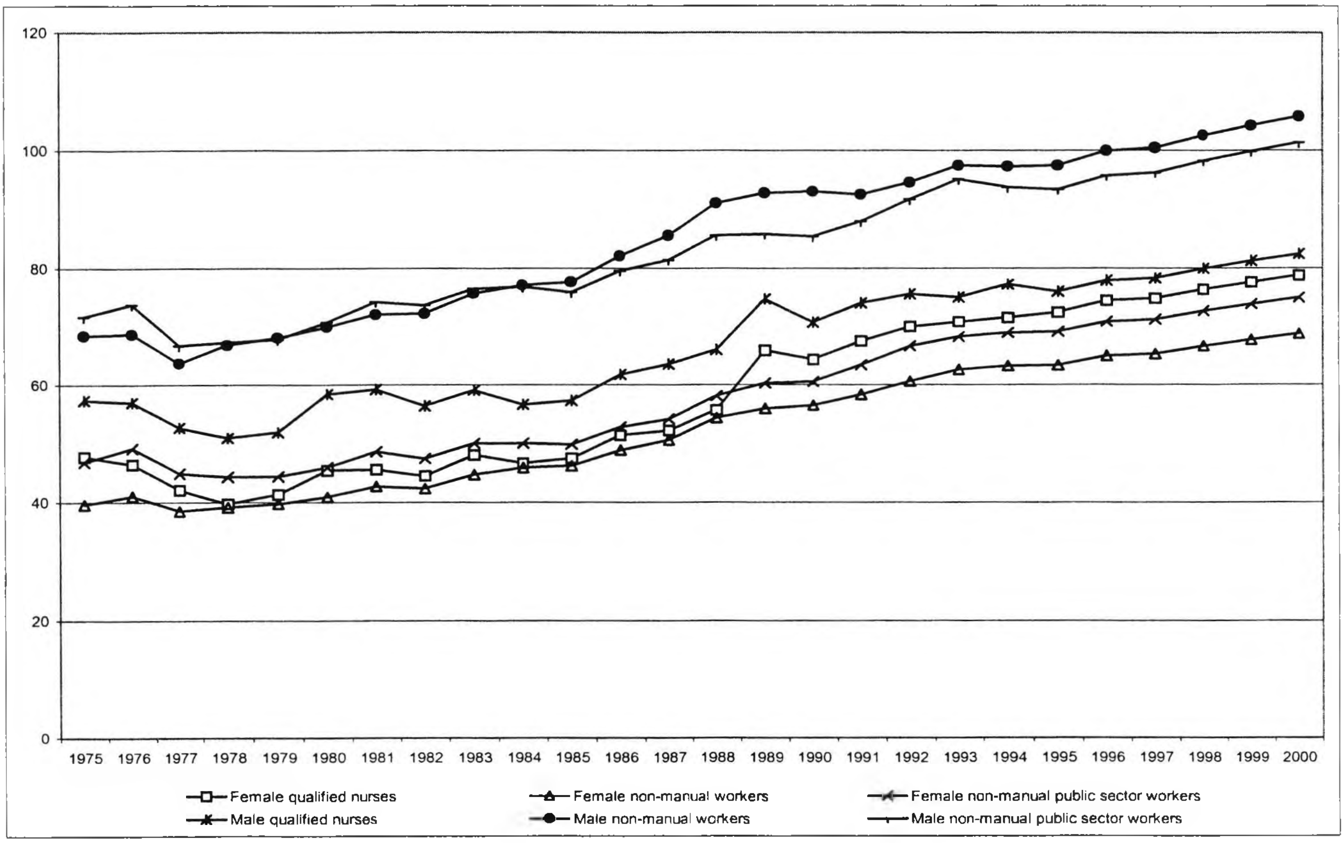


Figure 2. 1. Weekly earnings at constant 1975 prices, 1975-2000.

Figure 2.1 also illustrates trends in the mean earnings of non-manual workers and non-manual public sector workers (both of which include qualified nurses) across the period. The trends are generally upward, and are smoother than those for nurses, without the direct effects of fluctuations caused by the independent pay reviews. Note that while male nurses' earnings are generally greater than females nurses' earnings it is the case that male nurses earn substantially less than the comparator groups while female nurses earn more (at least, this is the case for females since 1989 following the clinical grading review). This has implications for the relative attractiveness of a career in nursing, which is an issue considered in more detail in subsequent chapters. The important point is that the vast majority of nurses (i.e. females) earn on average higher wages than workers in comparable occupations (i.e. other female non-manual workers). Note that since qualified nurses are included as non-manual workers then in reality the disparities are likely to be greater than those presented here. For example, for females if qualified nurses' earnings were not included in the estimates of non-manual workers' earnings then average non-manual workers' earnings would be depressed (qualified nurses' earnings are above the mean) and the positive earnings differential in favour of nurses would be even higher. The opposite is true for males.

More information on earnings relativities is presented in Figure 2.2. This figure highlights two main trends. First, as already mentioned male nurses earned less than other male non-manual workers across the period on average, and generally female nurses earned more than other female non-manual workers. Second, while there is some fluctuation the earnings differentials have generally persisted over time. Those years in which there was a marked improvement in the relative earnings of nurses correspond with those in which nurses' pay was reviewed independently. Thus relative earnings increased most noticeably in 1975, in 1980, and in 1989. Apart from these changes the earnings gap has remained fairly constant with female nurses earning roughly 5-10% higher wages than the other groups of female non-manual workers, at least since 1989, and male nurses earnings are roughly 20% lower than those in the comparator groups.

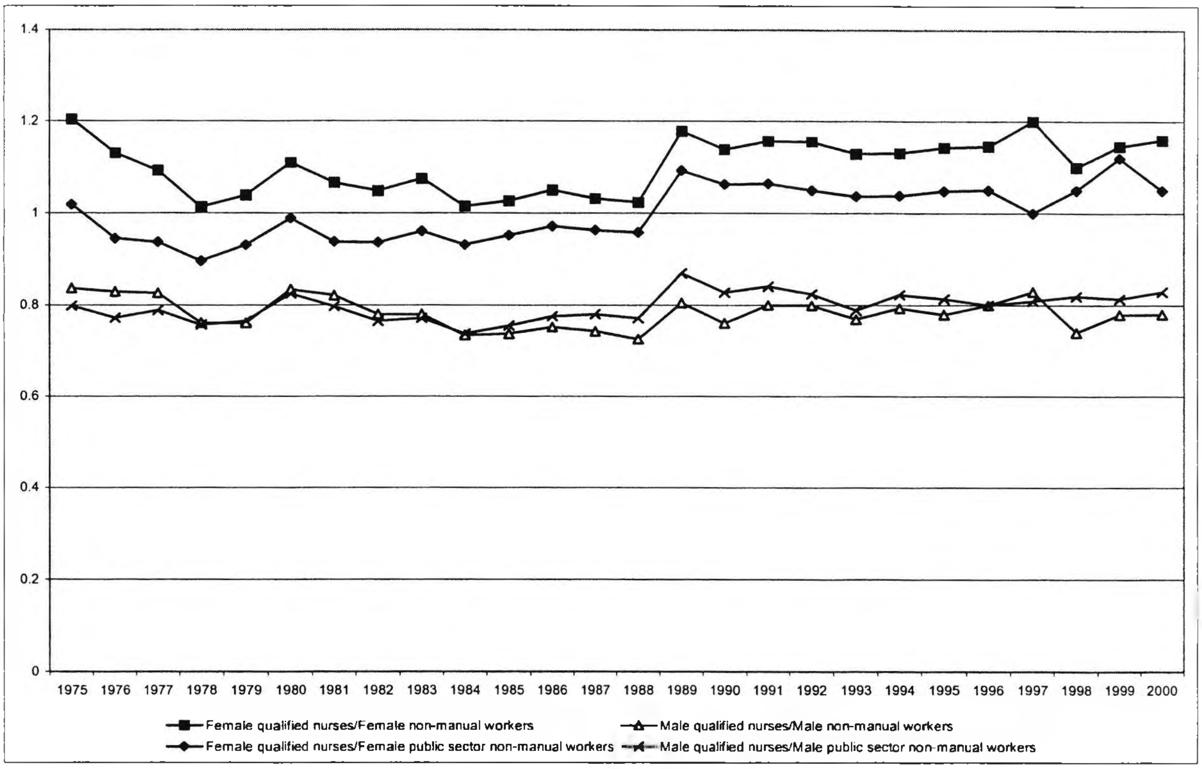


Figure 2.2. Ratio of earnings of qualified nurses to non-manual workers' earnings, 1975-2000.

In terms of these comparators note the differential for females between all non-manual workers and public sector workers. The implication is that female private sector non-manual workers earn higher wages than their public sector counterparts. For males there is generally no such disparity.

Growth in weekly earnings for a variety of occupation groups over the whole period 1975 to 2000 is presented in Figure 2.3. This indicates that within every occupation group females enjoyed greater increases than males. This is particularly true in the private sector (which is consistent with Figure 2.2) and is mainly due to the Equal Pay Act of 1975. Although equal pay between males and females had been achieved among almost all public sector non-manual workers prior to 1970, this was generally not true of the private sector. The equalisation brought about by the Equal Pay Act therefore explains why the largest increases over the entire period were achieved by female private sector workers. It is of note that all the female occupation groups considered obtained growth in real earnings in excess of 60% over the period 1975 to 2000. Only one male occupation group (male non-manual private sector workers) achieved a comparable increase. These findings are consistent with previous work in this area. For example, in a review of public sector earnings between 1970 and 1992 Elliott and Duffus (1995) concluded that nurses had enjoyed some of the largest increases in real earnings over this period and that they had also received the highest average size of wage settlement in the public sector.

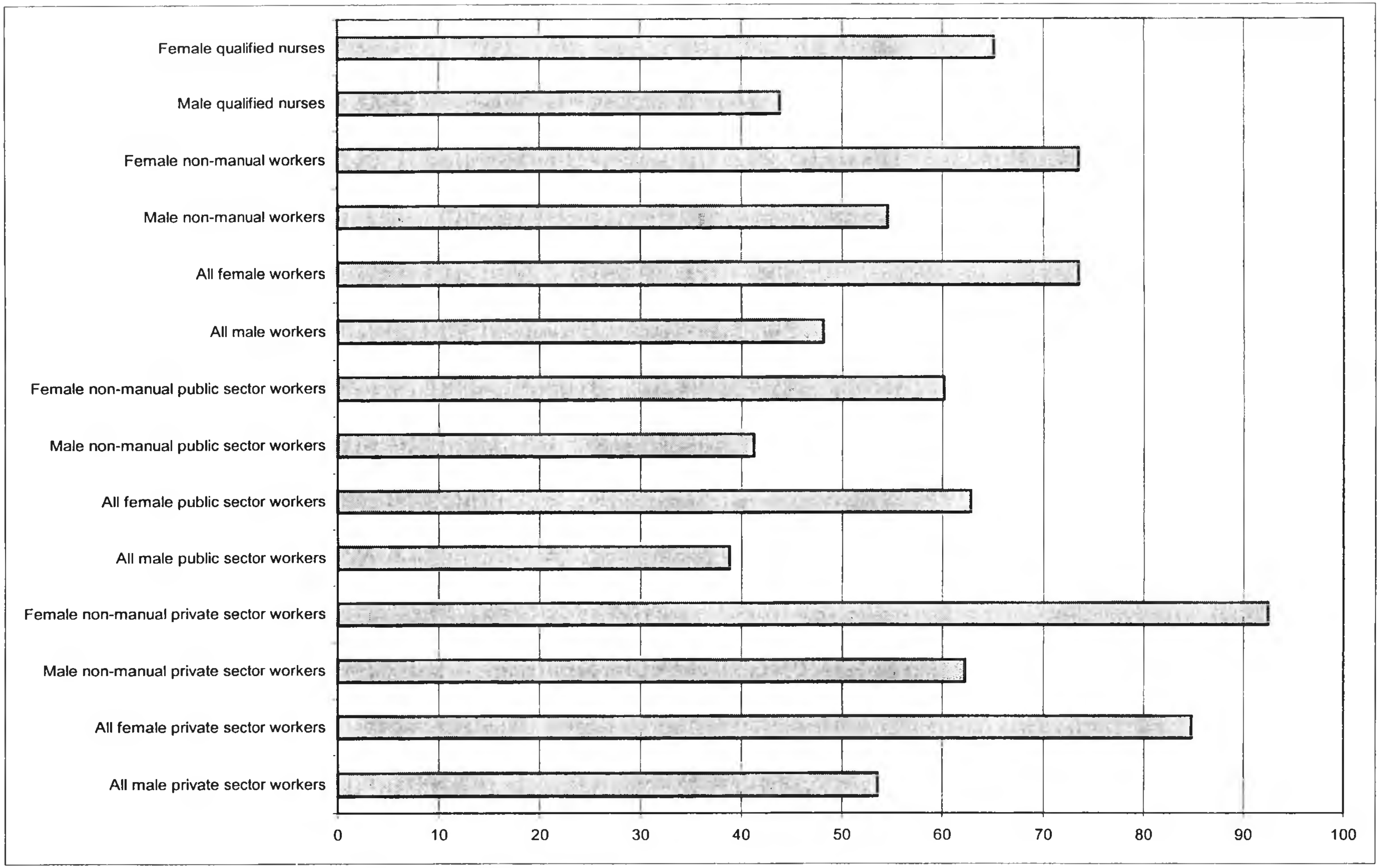


Figure 2.3. Growth in real earnings (%), 1975-2000

Table 2.1 presents various indicators of growth in mean earnings over the period 1975 to 2000. The trends within each time period are a combination of nominal changes in earnings and the rate of inflation. For example, from 1975 to 1980, the rate of inflation was high in Great Britain, and nurses generally only achieved small nominal increases in earnings following the substantial increase in pay awarded by the independent review in 1974. Conversely, in the period 1985 to 1990 nurses enjoyed substantial increases in real earnings largely because nominal earnings increased, particularly as a result of the clinical regrading exercise, but also because the inflation rate was falling.

	1975-1980	1980-1985	1985-1990	1990-1995	1995-2000	1975-2000	Annual mean
Female qualified nurses	-4.8	4.4	35.6	12.6	8.7	65.1	2.6
Male qualified nurses	1.8	-1.9	23.5	7.6	8.1	43.8	1.8
Female non-manual workers	3.3	12.9	22.2	11.9	8.8	73.5	2.9
Male non-manual workers	2.2	11.1	19.7	4.8	7.9	54.6	2.2
Female non-manual public sector workers	-1.9	8.5	21.5	14.0	8.3	60.2	2.4
Male non-manual public sector workers	-1.4	7.3	12.5	9.4	7.2	41.3	1.7
Female non-manual private sector workers	7.2	18.8	25.3	11.6	9.1	92.5	3.7
Male non-manual private sector workers	4.3	13.5	22.1	3.5	8.0	62.2	2.5

Table 2.1. Growth in real earnings, 1975-2000

Finally in this section we compare nurses' earnings with two crude measures of output. Figure 2.4 shows a comparison of changes in nurses' earnings with changes in the number of finished consultant episodes (FCEs) per nurse. This measure of output is calculated as the total number of hospital FCEs in England and Wales each year divided by the total number of whole-time equivalent nursing staff employed in NHS hospitals (qualified plus unqualified nurses and midwives). As discussed at length in Chapter 1 it is difficult to measure the contribution of nurses to the production of health care, and this output measure in particular is rather crude, especially in the context of qualified nurses. That said there does appear to be a positive relationship between the two variables, though this is unsurprising since the number of FCEs is likely to be determined by the magnitude of NHS expenditure. It should be stressed that this cannot be taken as evidence that nurses' earnings are linked to their output (in this case, their average physical product), though as discussed in Chapter 1 workload and productivity does feature in the Pay Review Body's deliberations over nurses' pay.

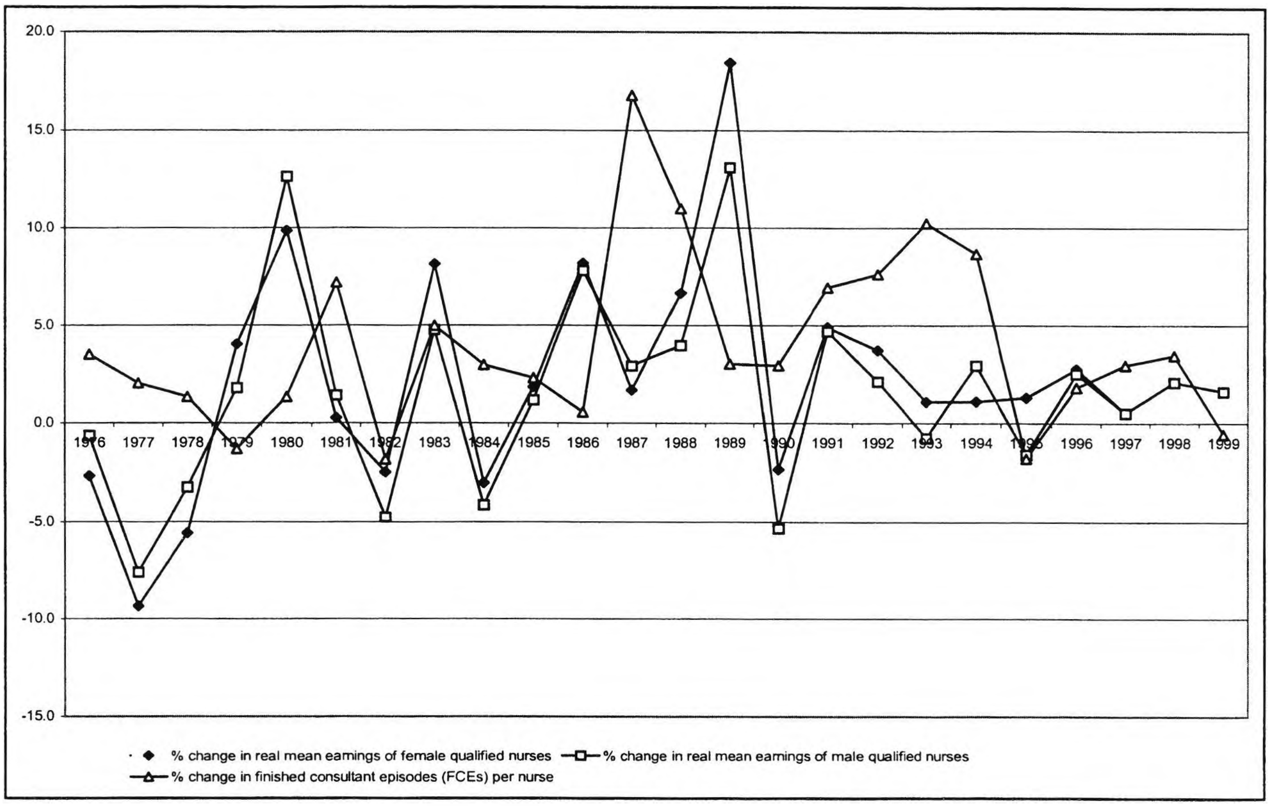


Figure 2.4. % Change in finished consultant episodes (FCEs) per nurse and % change in nurses' earnings, 1975-1999.

Finally, in Figure 2.5 we compare for England and Wales across the period 1975 to 1998 the number of available beds per nurse and nurses' real weekly earnings. Available beds per nurse are calculated as the total number of available hospital beds in England and Wales each year divided by the total number of whole-time equivalent nursing staff employed in NHS hospitals (qualified plus unqualified nurses and midwives). This shows that while nurses' earnings have generally increased there has been a steady reduction in the number of beds attended per nurse over the period. Rather than indicating a decline in the quantity of nursing output this is more indicative of a change in the nature of the output. Taken in conjunction with an increase over time in the numbers of patients treated (in Figure 2.4 the year-on-year changes in FCEs per nurse are generally positive), Figure 2.5 indicates that NHS throughput has increased (in simple terms, there are more patients and less beds). Presumably this is partly due to funding and partly due to advances in health care technology, meaning that procedures traditionally requiring an in-patient stay are now provided as day cases, on an outpatient basis or even in the community. The implication is that it is difficult to measure whether marginal productivity is increasing or decreasing because the type of nursing care provided (the nature of the output) is clearly evolving over time.

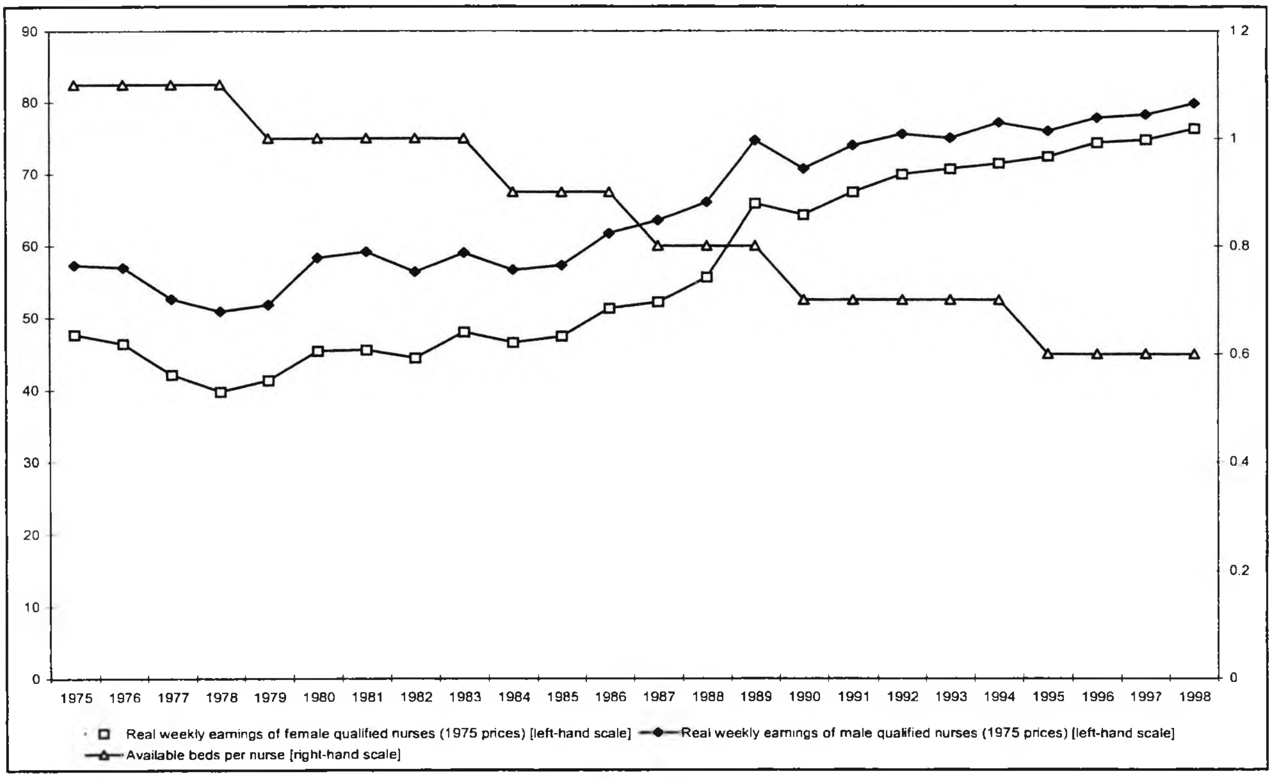


Figure 2.5. Available beds per nurse and nurses' real weekly earnings (1975 prices), 1975-1998.

2.3. Conclusions

In this chapter we have reviewed nurses' actual earnings illustrating the relationship between events in nurses' pay history and their earnings. We also presented a comparison of nurses' earnings with those of workers in other occupation groups. We found, for example, that female nurses (who comprise 90% of the workforce) earn on average higher wages than other female non-manual workers. These empirical findings inform the analysis of nurses' earnings in the next few chapters. In Chapters 4 to 6 we investigate the determinants of nurses' earnings and we examine whether the earnings differentials between nurses and other workers demonstrated in this chapter persist because of nurses' superior human capital characteristics, or whether there is a premium to being employed as a nurse.

Before considering this issue we continue our analysis of nurses' earnings by examining the financial costs and benefits to being employed as a nurse. This helps us to understand why an individual might choose a career in nursing. Rather than focusing on mean earnings at one point in time we instead look at the present value of lifetime earnings. This takes into account the costs and benefits of choosing to be employed as a nurse over the lifetime and helps us to understand the human capital effects of nursing which will be the subject of future chapters. In Chapter 3 we therefore measure the private net present value and private internal rate of return to being employed as a nurse in Great Britain.

CHAPTER 3

THE PRIVATE NET PRESENT VALUE AND PRIVATE INTERNAL RATE OF RETURN TO BECOMING A NURSE IN GREAT BRITAIN

3.1. Introduction

Given the vacancy rates in the British nursing labour market and the associated recruitment problems discussed earlier as well as the on-going discussions on the relative earnings of nurses further consideration of the financial return to becoming a nurse is warranted. Two related measures – the net present value (NPV) and the internal rate of return (IRR) – provide summaries of the returns to human capital investments and in this way may be used to determine the attractiveness financially of a career in nursing in Great Britain. In the context of the nursing labour market, NPV and IRR analyses are based on the assumption that the training and qualifications required to become a nurse (a pre-requisite for entry into the profession) may be thought of as an investment in human capital. This investment involves incurring a cost during the period of its acquisition but yields benefits in the form of improved earnings capacity for the educated individual at a later stage.

A key feature of NPV and IRR analyses is that while they measure the costs and benefits of investments in human capital they tend to concentrate on quantifiable economic costs and benefits, including the financial cost of training and education and the subsequent financial earnings of the individual. They tend to ignore non-wage factors associated with investments in human capital. These might include non-wage advantages and disadvantages associated with different occupations, consumption costs and benefits directly associated with training and education programmes, and non-employment benefits of training and education. If an individual associates an investment in human capital with non-wage advantages or disadvantages then the net financial benefit required to attract an individual into that investment will be different than if the non-wage factors were negligible or could be ignored.

The literature on using NPVs and IRRs as measures of the attractiveness of investments in human capital is huge. These types of analyses have most frequently been conducted to assess the returns to non-compulsory schooling or higher education and have usually examined the returns to education in broad bands such as all first university degrees or all postgraduate degrees (Maglen and Layard, 1970; Metcalf, 1973; Morris and Ziderman, 1971; Pissarides, 1982; Psacharopoulos, 1972, 1973, 1981, 1985; Psacharopoulos and Layard, 1979; Ziderman, 1973). Some analyses have considered particular professions, including physicians (Birch and Calvert, 1973; Burstein and Cromwell, 1985; Mott and Kreling, 1994; Wilkinson, 1966; Wilson, 1980, 1983a, 1983b, 1984, 1985a, 1985b, 1987a, 1987b). Although some are dated, all of these studies found declining private internal rates of return to professional occupations over time. There has been no comparable study of the returns to nursing in Great Britain. That is the primary aim of this chapter.

3.2. General methodology

Two types of NPV and IRR may be calculated – private and social. The private NPV and the private IRR measure how attractive a particular investment in human capital is to the individual. They relate to the demand for places on training and education programmes, and are an indicator of whether the future demand for places is likely to rise or fall as entry into the profession adjusts in response to perceived returns. Because they model individual decisions the private NPV and private IRR are helpful in explaining vacancy rates in labour markets. Alternative measures are the social NPV and the social IRR, which assess the attractiveness of an investment in human capital to society. The social NPV and the social IRR may be used as indicators of the efficiency of existing public funding of education and training. In calculating these measures earnings gross of taxation are usually used as a proxy for net contributions to output. The underlying assumption is that the wage rate equals the MVP. For the reasons outlined in Chapter 1 this is entirely unrealistic in the nursing labour market in Great Britain. First, because it is exceedingly problematic to measure the MVP of employing an additional nurse. Second, even if this were possible wages are not determined in a perfectly competitive environment, but are instead determined administratively by the Pay Review Body which is heavily influenced by the size of the budget allocated to the NHS and the monopsony power of employers. For these reasons and because we are primarily interested in modelling individual decisions to enter the nursing labour market we concentrate our analysis on the private NPV and the private IRR. We do not consider the social returns further.

The private NPV to an investment in human capital is calculated by solving the following:

$$\mathrm{NPV} = \sum_{t=18}^{60} \frac{B_t - C_t}{(1+r)^{t-17}} \qquad [3.1]$$

B_t is the financial benefit to the individual of undertaking the investment at age t, and C_t is the financial cost, and r is the rate at which the individual discounts the future (the individual's marginal time preference rate, MTPR). The investment is assumed to begin at age 18 and retirement is assumed to occur at age 60. The benefits of the investment are the grant or bursary received by the individual while undergoing training and the earnings of the individual subsequent to training net of taxation. The costs of the investment are the earnings foregone by the individual while undergoing training and subsequent to it (that is, the earnings the individual would have received had they not made the investment). Faced with a single possible investment in human capital the NPV criterion dictates that the individual should undertake the investment if the NPV exceeds zero. Where the choice is between different investments in human capital and these investments are mutually exclusive the general rule is to select the project with the highest NPV.

The private IRR is calculated by solving for r in the following expression:

$$\sum_{t=18}^{60} \frac{B_t - C_t}{(1+r)^{t-17}} = 0 \qquad [3.2]$$

Faced with a single possible investment in human capital the IRR approach is to calculate the IRR and compare it to the individual's MTPR. The rule for undertaking the investment is to accept if the IRR is greater than the MTPR (the implication is that the NPV must then be greater than zero). When comparison of mutually exclusive investments is required it is not necessarily the case that the investment with the highest IRR should be preferred and the NPV approach should generally be preferred.

The commonly used description of NPV and IRR analyses as 'rate of return' studies or as measuring the 'returns' to investments in human capital reflects the fact that most studies of this kind involve only the calculation of IRRs. Some do compute NPVs however (e.g. Lindsay, 1973, 1976; Wilson, 1987; Wilkinson, 1966; Morris and Ziderman, 1971; Siebert, 1977). Unfortunately both the NPV and the IRR methods have their weaknesses and they may yield conflicting results concerning the attractiveness of an investment in human capital. Generally the NPV approach is considered to be less flawed (see, for example, Thompson, 1980; Sugden and Williams, 1978; and, Pearce and Nash, 1981, for useful discussions). Conceptual difficulties with the NPV and IRR approaches are discussed in detail in Appendix 3.1. The main points for discussion concern the existence of mutually exclusive investment opportunities, the potential absence of a unique IRR, the fact that both the IRR and the NPV ignore the time pattern of costs and benefits, and the possibility that the MTPR may vary over the lifetime of the project. The general conclusion is that both the IRR rule and the NPV rule have a number of potentially serious weaknesses as summary measures of the attractiveness of investments in human capital. However, because of its superiority in comparing mutually exclusive investment opportunities on balance the NPV has fewer weaknesses and should be preferred. Nonetheless, the relative ease of calculation of the IRR criterion means that estimation of this measure is included in the present analysis.

Data constraints are the main limitations of NPV and IRR analyses. As the majority of nurses (some 90%) in Great Britain are female we concentrate on the returns for females only. For comparison, we also calculate the private NPV and private IRR to females to becoming teachers and to obtaining a university degree. These may be thought of as realistic alternatives for individuals who choose to become nurses, and their inclusion provides a more realistic scenario in terms of occupation choice where individuals choose between two or more mutually exclusive projects. It should be borne in mind that teachers generally and some nurses are graduates and therefore the returns to obtaining a university degree estimated for graduates may in part include the returns of workers also included in the other two options.

It is usual to think of investments in human capital as investments in training and education programmes which augment the individual's stock of human capital. One important issue then is whether it is meaningful at all to compute NPVs and IRRs for particular occupation groups (e.g. nurses and teachers, as considered here) rather than for particular courses of education and training (e.g. obtaining a nursing qualification or undergoing teacher training). As noted above returns have been estimated for different occupation groups in the literature (see, for example, Birch and Calvert, 1973; Burstein and Cromwell, 1985; Mott and Kreling, 1994; Wilkinson, 1966; Wilson, 1980, 1983a, 1983b, 1984, 1985a, 1985b, 1987a, 1987b). When an investment in education and training is defined as being occupation specific and a pre-requisite for entry into an occupation then the IRR and NPV for that investment may be viewed as in fact pertaining to the occupation. This is the case for nurses in the NHS. However, occupational status is not fixed over time, and individuals who obtain a nursing qualification are not obliged to stay in that profession for the rest of their working lives. Because we are interested primarily in explaining the current state of play in the nursing profession (the occupation) in the present analysis we concentrate on the occupation group rather than on the education and training group. We estimate the NPV and IRR to being employed as a nurse (for which obtaining a nursing qualification is currently a pre-requisite) rather than for undergoing nurse training (completion of which does not compel the individual to work as a nurse throughout their working lives, and estimation of which will include earnings profiles for non-nurses).

The private NPV and private IRR are calculated for the years 1991 to 1996. Partly this time frame was chosen due to data constraints. However, it is the case that returns across this time period are not directly comparable with those of earlier periods due to a number of wide-ranging changes both to the nursing profession in the NHS and to the NHS as a whole. As noted in Chapter 2 the late 1980s saw major changes to the structure of the nursing profession via the clinical regrading exercise and the pricing of this new structure by the Pay Review

Body. Additionally there were significant changes to pre-registration training for nurses with the introduction of the Project 2000 reforms. These changes, along with the introduction of the NHS Reforms in 1991, mean that NPV and IRR estimates pre-1991 may reflect structural changes to the nursing profession and the NHS in addition to changes in pay levels.

3.3. The data

Central to NPV and IRR calculations are age-earnings profiles that depict earnings at each year of age for individuals with particular levels and types of education and training. A hypothetical age-earnings profile is presented in Figure 3.1. For an individual undergoing a three-year period of training at age 18 with subsequent employment the age-earnings profile is depicted by the area 0abdef. 0acg depicts the age-earnings profile of the same individual not undergoing training but instead joining the workforce at age 18. Undertaking the investment incurs a (net) cost of magnitude bcde while undergoing the period of training and a (net) benefit of efg throughout the lifetime of the individual. In intuitive terms the NPV method measures whether efg – bcde is positive or negative at a specified MTPR. The IRR method seeks to find the MTPR at which efg = bcde.

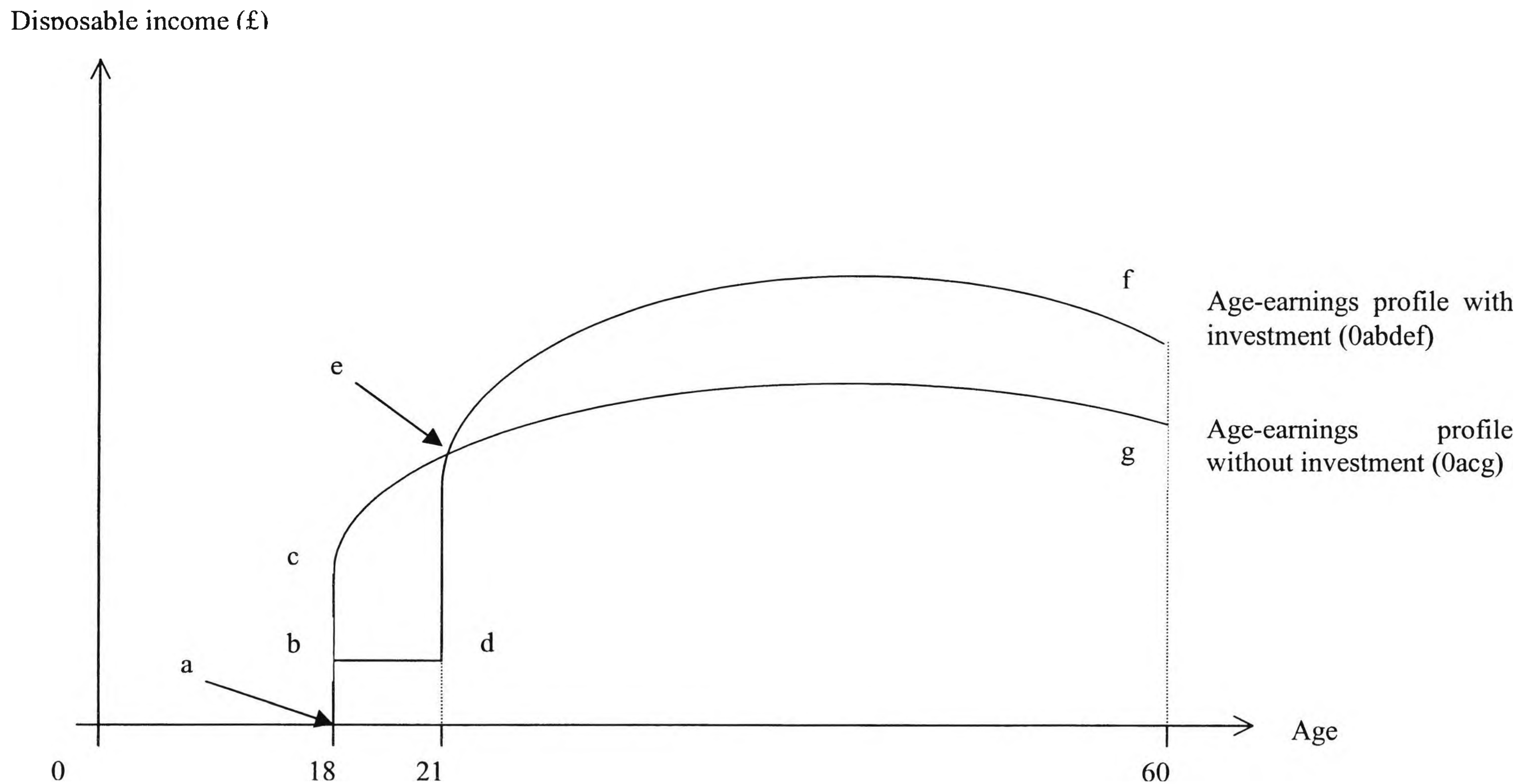


Figure 3.1. Hypothetical age-earnings profiles used to calculate the NPV and IRR for investments in human capital

There are a number of practical difficulties that arise in estimating NPVs and IRRs using the methods described so far. These arise first because simple comparison of basic earnings data by age for different occupation groups is unlikely to result in accurate estimation of the true costs and benefits of investments in human capital. The second type of difficulty arises due to the nature of the data available for NPV and IRR calculations. We discuss these issues in further detail in Appendix 3.2. Of particular importance are the potential conflation of age and cohort effects in the use of cross-section data, the effects of comparative advantage and ability bias, non-inclusion of the non-wage costs and benefits, and the effects of investments in human capital on work hours. Of these, one important issue worth re-emphasising here is that in terms of the data required for the basic age-earnings profiles the method adopted in this study is to use cross-sectional data on mean earnings of individuals at each age and in each occupation/qualification group. This is usual practice in NPV and IRR studies. Unfortunately this may lead to a conflation of age and cohort effects. However calculating the NPV and IRR in this way is justifiable for the following reasons. First in order to calculate the private NPV and private IRR to nursing in 1991 to 1996 we would ideally know the value of future earnings at older ages. For example, we wish to know the future lifetime earnings of an 18 year old nurse who enters the profession in 1996. These are obviously unobservable at the present time. In the context of modelling decisions of occupational choice an individual deciding whether to enter into the nursing profession is unable to ascertain what their true future earnings will be. While future earnings are unobservable what is observable is data on current earnings at older ages. It seems reasonable to suppose that individuals will use current earnings at older ages in an occupation as a predictor of future earnings at older ages. Second, from Equations [3.1] and [3.2] future costs and benefits are discounted anyway in the NPV and IRR calculations. Therefore, the importance of future unobservable earnings differentials diminishes with time. Possible divergence between unobservable earnings at older ages in future years and currently observable earnings at older ages becomes less important to the individual's choice of occupation.

We present some tentative comparisons concerning the potential conflation of age and cohort effects in Figure 3.2. We compare the mean earnings of nurses in 1980 at ages 18 to 34 years (a sixteen year age range based on earnings data in a single calendar year = cross-section data) to mean earnings of nurses aged 18 to 34 years from 1980 to 1996, respectively (i.e. earnings of nurses aged 18 years in 1980, aged 19 years in 1981 and so on = cohort data). 1980 is chosen as the index year because this is the first year that age-earnings profiles by occupation are available for the New Earnings Survey. This is therefore the longest time period obtainable for comparing the age and cohort effects. The general finding is that earnings of nurses at older ages in 1980 underestimate future earnings at older ages across the period 1980 to 1996. For example, relative to 18-year-old nurses in 1980 34-year-old nurses in 1980 earn higher wages (£85 per week compared to £60 per week ≈ 40% higher wages). However the earnings of 34-year-old nurses in 1996 (i.e. what an 18 year old nurse in 1980 would earn in 1996 if she remained in the profession) are even greater (£140 per week versus £60 per week ≈ 133% higher wages). This suggests that in 1980 current earnings data on nurses at older ages are a relatively poor predictor of future earnings at older ages. In terms of the present analysis it is very important to bear in mind that it is not possible to ascertain whether this trend applies to the years 1991 to 1996 (the index years in the present analysis), precisely because earnings in future years are unobservable. Further, to repeat the arguments of above, individuals in the period 1991 to 1996 deciding whether to enter into the nursing profession are unable to ascertain what their true future earnings will be. The assumption in the present analysis is that because future earnings are unobservable individuals will use current earnings at older ages in an occupation as a predictor of future earnings at older ages.

Also, future costs and benefits are discounted anyway in the NPV and IRR calculations so that the importance of future unobservable earnings differentials diminishes with time. Additionally, it is worth bearing in mind that while for example using 1980 earnings data on earnings at older ages underestimates future earnings at older ages it is unclear what effect this will have on the NPV and IRR calculations because the emphasis is not on absolute levels of earnings but on earnings differentials across occupation groups.

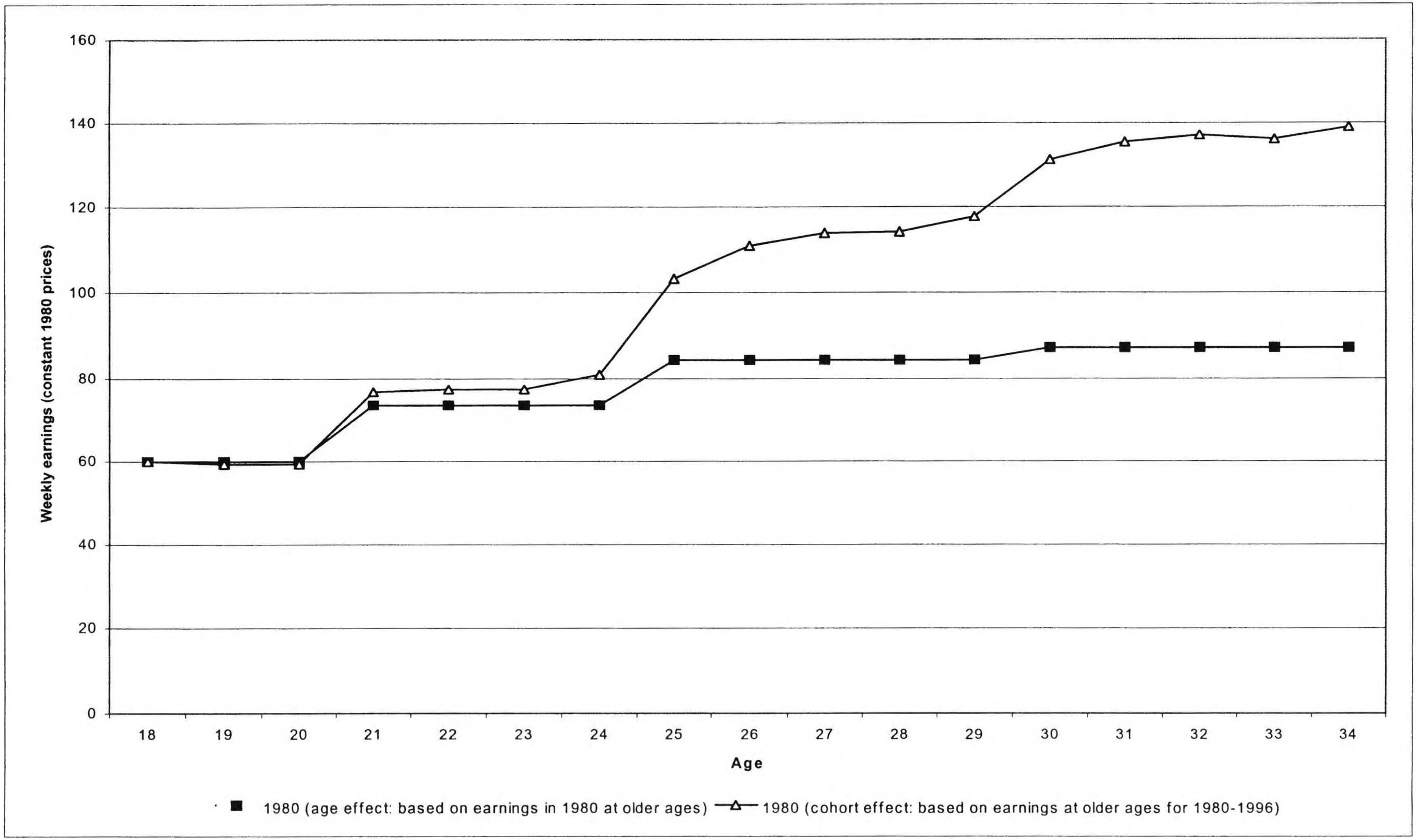


Figure 3.2. Comparison of nurses' earnings in 1980 by age to earnings in 1980-1996 by age (source: New Earnings Survey, selected years)

Age-earnings profiles for nurses and teachers were obtained from the New Earnings Survey (NES), an annual survey covering 1% of all employees in employment in Great Britain. [10] Note that the readily available published results of the NES do not include earnings by occupation by age (they include earnings by occupation or earnings by age) and therefore the age-earnings profiles for nurses and teachers were obtained via a special query on the NES dataset by the Office for National Statistics who hold the data. Earnings for individuals whose highest academic qualification was a degree were obtained from the British Household Panel Survey (BHPS), an annual longitudinal survey of over 17,000 individual household members living in Great Britain. All earnings data are for full-time workers and are calculated in five-year bands.

An important component of the cost of undertaking an investment in human capital is the opportunity cost earnings of the individual who undergoes training. In Figure 3.1 the opportunity cost earnings are 0acg. The ideal opportunity cost age-earnings profile would depict the earnings the individual would have received had they not made the investment. Because nurses and teachers and many of the occupations that individuals with degrees will choose to enter on completion of their investment are classed as being in the non-manual group, the earnings of female non-manual workers are used to represent the baseline opportunity cost earnings in the analysis. The assumption is that non-manual workers and nurses are comparable in terms of their labour market and personal characteristics. [11]

As discussed in Appendix 3.2 individuals who undertake an investment in their human capital and then enter a specific occupation may have a comparative advantage in terms of natural

[10] In the NES it is unfortunately not possible to distinguish between nurses working in the NHS and private sector nurses. Therefore the analysis considers all nurses (public sector and private sector) together. The inclusion of private sector nurses is unlikely to have a major impact however since the majority of nurses (some 85%) are employed by the NHS.

ability and motivation in that occupation. If this is the case then the earnings of non-manual workers may not accurately reflect the opportunity cost earnings of undertaking nurse or teacher training or obtaining a degree: the mean earnings of other non-manual workers may overestimate the earnings a nurse would receive if they were employed elsewhere. For this reason we also calculate the NPV and IRR using earnings of all female workers and female workers whose highest academic qualifications are A levels as opportunity cost earnings. Opportunity cost earnings for female non-manual workers and all female workers were obtained from the NES. Those for females whose highest academic qualifications were A level were obtained from the BHPS. Note that qualified nurses and teachers (and probably graduates) are included in the non-manual workers group. It is not possible to remove this effect (for example by removing nurses and teachers from the comparison group) because there is no information on the proportion of non-manual workers in each age group in the NES sample who are nurses/teachers. This is because the data used are not the typical NES published data (data on earnings by occupation by age are not usually published) but are based on a special query of the NES dataset by the Office of National Statistics. Unfortunately precisely for this reason it is not possible to ascertain the magnitude of effect on the results of including nurses and teachers in the figures for non-manual workers.

Mean earnings data across all ages for females working in the occupations groups considered are presented in Figure 3.3. The data for nurses, teachers, all workers and non-manual workers were obtained from the NES. This measures average weekly earnings for full-time female workers. The data for graduates and workers whose highest qualification was A level were obtained from the BHPS. These data were for average weekly earnings for female workers who worked in excess of 30 hours per week (working 30 hours per week or more is

[11] We consider this issue in further detail in Chapters 5 and 6.

used to define full-time workers in the BHPS). The rank ordering in descending order of magnitude by mean weekly earnings is the same in every year: teachers, then graduates, then nurses, then non-manual workers (which includes teachers, nurses and probably graduates), then all workers, and finally workers whose highest qualifications are A level. On the basis of this ranking one possibility is that the private NPV for nursing, teaching and obtaining a degree will be ranked in this same order relative to the opportunity cost earnings profiles. However, this may not necessarily be the case since we are concerned not with mean earnings across all ages but earnings at each age of the age-earnings profile. It does not necessarily follow that the rank ordering at each age will be the same as the rank ordering across all ages. We explore this issue in greater detail below by examining the actual age-earnings profiles in every year.

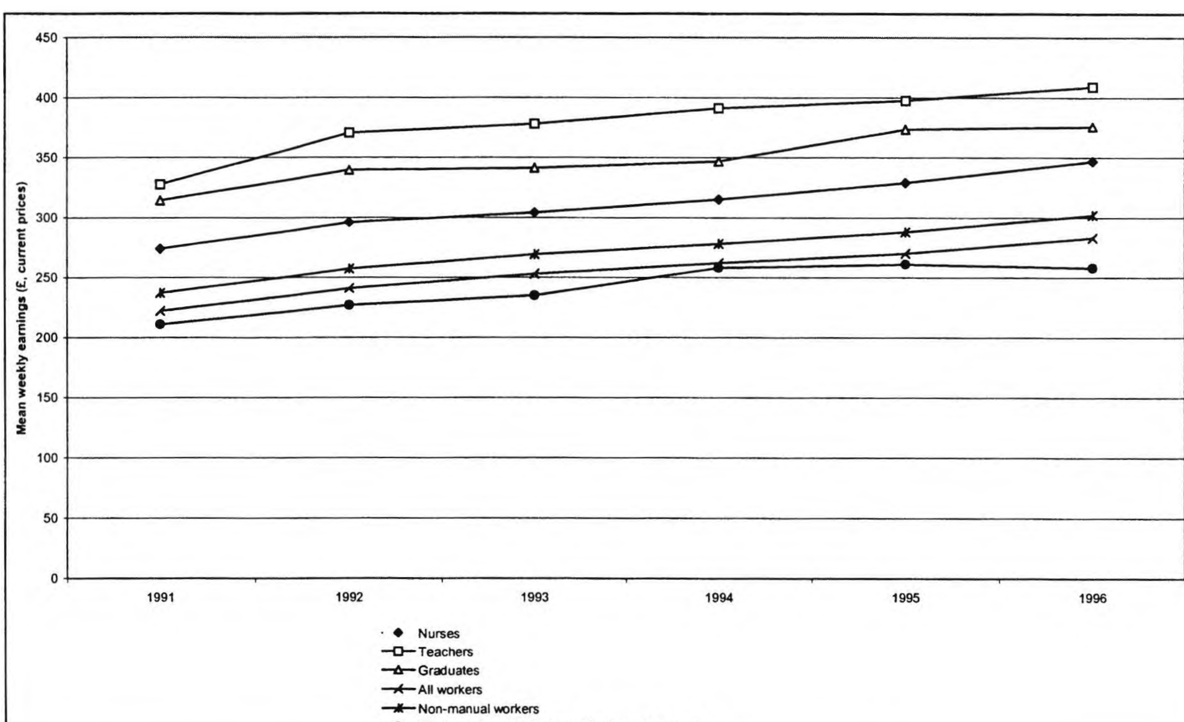


Figure 3.3. Mean weekly earnings in selected occupations, 1991-1996. Source: NES and BHPS

The length of nurse training and the length of time to obtain a degree are assumed to be three years. The length of training for teachers is assumed to be four years (either by a four-year teaching degree or via a three-year non-teaching degree followed by a one-year postgraduate diploma). The direct costs of nurse training were taken from Netten et al. (1998) and deflated to prices for the relevant years using the NHS pay and prices index. The direct costs of initial teacher training and obtaining a degree were taken from the Higher Education Statistics for England (Department of Education, selected years).

The bursary received by individuals training to be nurses was taken from Netten et al. (1998) and deflated to relevant year's prices using the NHS pay and prices index. The grant received by individuals undergoing initial teacher training and university education was taken from the Higher Education Statistics for England (Department of Education, selected years).

Earnings net of taxation (income tax and national insurance) are used to calculate the private NPV and private IRR. Income tax allowances, rates of income tax, and national insurance thresholds and rates used to calculate earnings net of taxation were taken from Annual Abstract of Statistics (ONS, selected years). For the income tax adjustments it is assumed that the entire married couple's allowance was allocated to the husband.

While the data described above may be used to construct basic age-earnings profiles, these data on their own may not provide a complete picture of the net benefits of investments in human capital (see Appendix 3.2). Consequently we adjust the data to obtain a more accurate picture of the true costs and benefits. Four adjustments are made: for mortality; for unemployment; for other causes of economic inactivity; and, for discontinuation from training.

It is common in NPV and IRR analyses to make some adjustment to the age-earnings profiles for mortality. The approach usually adopted and the approach adopted here is to multiply the earnings at each age by the probability that the individual will survive to that age. It is assumed that mortality rates do not differ across occupation groups. Mortality rates were taken from English life tables (Government Actuary Department, 1992).

The returns to training are likely to be affected positively by the higher employment rates of more educated individuals. Indeed, improved employment rates may be one reason why individuals choose to undertake training in the first place. Therefore, some adjustment to age-earnings profiles is justified for employment. Earnings at each age are multiplied by the probability that individuals of that age are employed. Employment rates for each occupation group were taken from the BHPS.

Non-participation in the workforce for reasons other than unemployment (retirement, family care, long term sickness or disability and maternity leave) may also affect the NPV and IRR. The approach adopted here is to multiply the earnings at each age by the probability that the individual will participate in the workforce at that age. It is assumed that participation rates for reasons other than unemployment do not differ across occupation groups. Participation rates were taken from the BHPS.

Whilst an individual may decide to undertake training there is no guarantee that they will successfully complete that training. Some allowance would therefore seem to be justified for the possibility of failure to complete the course. We assume here that an individual who drops out earns a zero return on the investment and achieves the earnings profile they would have achieved had they not begun training in the first place. Course discontinuation rates for nurse training were taken from the ENB Annual Report (English National Board, selected years). Dropout rates for initial teacher training were calculated from the Annual Abstract of Statistics (ONS, selected years). Dropout rates for degree courses were taken from the University Management Statistics and Performance Indicators in the UK (Committee of Vice Chancellors and Principals and Universities Funding Council, selected years).

Workers in some occupation groups work longer hours than others. For these individuals the NPV and IRR methods may involve an overstatement of the returns to an investment in human capital if they fail to take account of the longer hours worked (Lindsay, 1971). The theoretical argument is that the use of human capital to earn an income involves some disutility in the form of leisure time foregone. Therefore the measured return reflects, partly at least, a return to working longer hours and not an excess return to investment. One option often employed in previous research (see for example, Eckhaus, 1973, Lindsay, 1973, 1976) is to standardise earnings profiles at each age by the number of hours worked.

The main problem with including any adjustment of this kind lies in the issue of obtaining comparable measures of hours worked in different occupations. In Table 3.1 we present hours worked for different occupation groups considered in the present analysis for the period 1991 to 1996. The data for nurses, teachers, all workers and non-manual workers were obtained from the New Earnings Survey. This measures total average weekly hours worked per week (normal basic hours plus overtime hours) for full-time female workers. The data for graduates and workers whose highest qualification was A level were obtained from the British Household Panel Survey. These data were for total average hours worked per week (including overtime) for female workers who worked in excess of 30 hours per week (working 30 hours per week or more is used to define full-time workers in the survey).

Year	Nurses [1]	Teachers [1]	Graduates [2]	All workers [1]	Workers whose highest qualification is A level [2]	Non-manual workers [1]
1991	37.6	29.1	39.1	37.4	33.7	36.8
1992	37.7	28.7	38.1	37.3	33.2	36.8
1993	37.5	29.7	38.7	37.4	33.6	36.9
1994	37.6	30.1	39.9	37.6	33.5	37.0
1995	37.8	30.0	39.3	37.6	34.6	37.0
1996	37.8	30.2	40.6	37.6	34.8	37.1

[1] Data obtained from the New Earnings Survey
[2] Data obtained from the British Household Panel Survey

Table 3.1. Total weekly hours worked on average by female workers in different occupation groups

Looking at nurses, graduates, non-manual workers (which includes nurses and probably graduates) and all workers first we can see that generally most workers in these groups worked on average between 37 and 40 hours per week across the time period considered. It is interesting to note that on average non-manual workers worked shorter hours than all workers implying that manual workers work longer hours on average than non-manual workers. Workers whose highest qualification is A level worked on average less hours than the other occupation groups (33 to 35 hours per week), with the exception of teachers for whom the average hours worked is even less again. The apparently very low hours of work for teachers (mean 29 to 30 hours per week) demonstrates some of the pitfalls inherent in obtaining comparable measures of hours worked in different occupations. These low figures are a reflection primarily of the number of contracted hours that a typical teacher is expected to be in attendance at school. Hours worked after school hours, organising out-of-school or after-school activities, meeting with parents, preparation time, marking time and time spent writing reports are not included. Clearly, a more comprehensive analysis taking these factors into account would most likely show that teachers generally work considerably longer hours in total that the data in this table suggest.

Caveats concerning the comparability of data notwithstanding it would appear that some adjustment to the NPVs and IRRs estimated in this paper is justified, at least as a sensitivity analysis. We therefore re-estimate the calculated NPVs and IRRs adjusting for the hours worked in the different occupation groups. For the purposes of the calculation we adjust all earnings profiles to a standard 40-hour working week: earnings at every age in each occupation group are multiplied by a factor of 40/h where h is the mean hours worked per week in that occupation group as shown in Table 3.1.

3.4. The private NPV and private IRR to becoming a nurse in Great Britain

Figure 3.4 plots age-earnings profiles used to calculate the private NPV and private IRR in 1996. Trends in age-earnings profiles over the period 1991 to 1995 were similar (see Appendix 3.3). From age 18 to 30 nurses' disposable income was generally the same or greater than that of teachers and graduates (in some years – see Appendix 3.3 – it was slightly lower). From age 30, it was generally lower, except after age 55 when nurses' disposable income was greater than that of graduates. Apart from during the initial period when education and training took place nurses', teachers' and graduates' disposable income was greater than that of non-manual workers, all workers and workers whose highest qualifications were A levels. Note that the rank ordering of occupation groups by mean weekly earnings across all ages shown in Figure 3.3 is not reflected across each age in the age-earnings profile. The effect of the discounting process means that this rank ordering will not necessarily be reflected in the ranking by NPV.

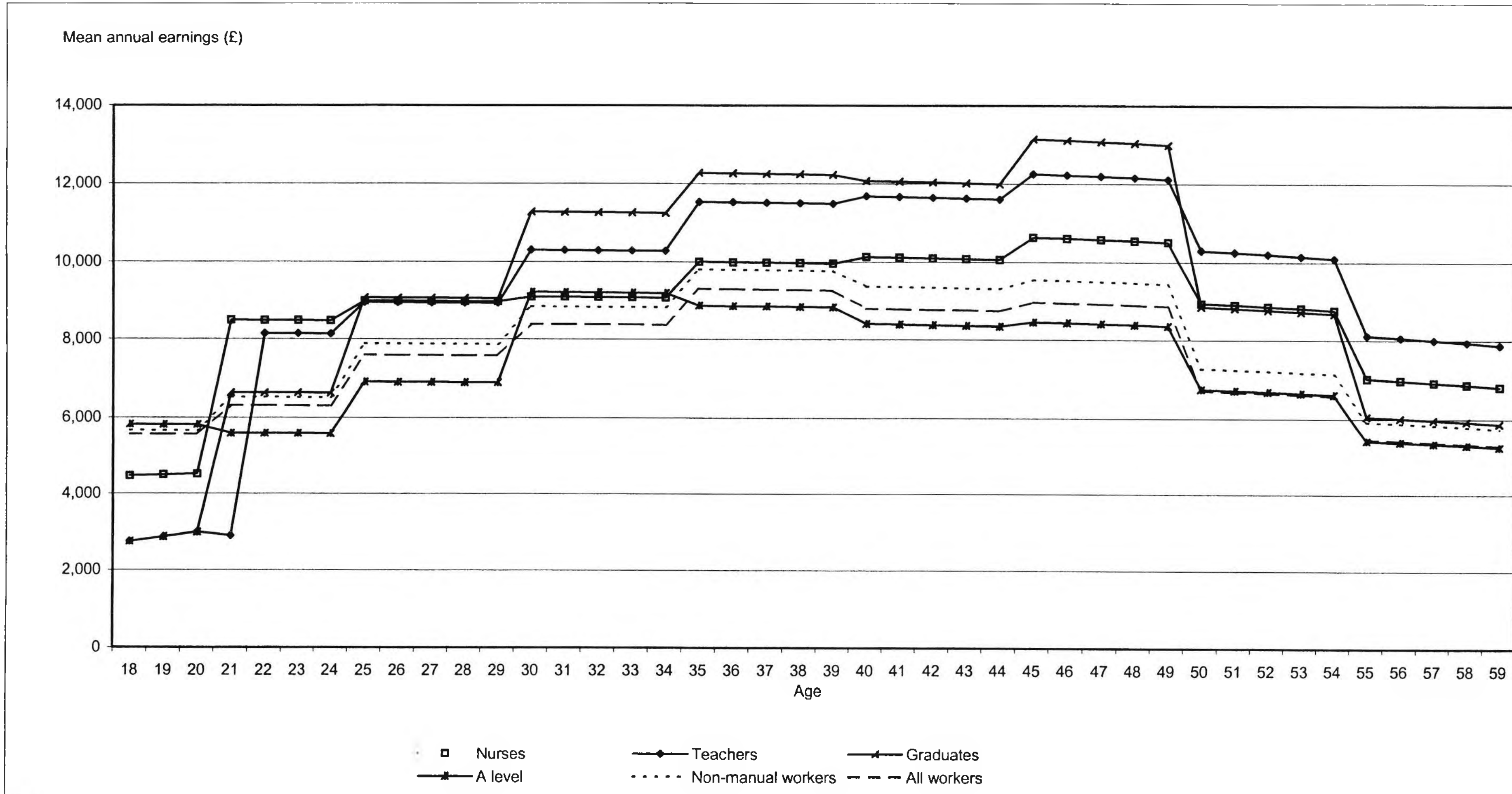


Figure 3.4. Actual age-earnings profiles used to calculate the private NPV and private IRR for 1996 (earnings net of taxation [income tax plus national insurance] adjusted for mortality, unemployment, other causes of economic inactivity, and discontinuation from training). Source: see text.

A summary of results using the NPV rule is given in Tables 3.1 to 3.3, which shows which of the three investments (becoming a nurse, teacher or graduate) yields the highest NPV at different MTPRs (from 0% to 40%). Table 3.1 presents the NPV results estimated using the earnings of non-manual workers as the opportunity cost investment (the baseline analysis). Tables 3.2 and 3.3 show results estimated using all workers and workers whose highest qualification is A level, respectively, as the opportunity cost investments. From Table 3.1 we can see that taking non-manual workers as the opportunity cost investment in 1996 for individuals with an MTPR of 0%-1% teacher training would be the preferred investment since it yields the highest NPV at those MTPRs. For individuals with an MTPR of 2%-7% obtaining a degree would be preferred, and for individuals with an MTPR of 9%-34% nurse training would be preferred. Across each of the years considered nurse training yields the highest NPV if the individual's MTPR is 8%-12% or more. This difference across years is a function of small differences in the relative age-earnings profiles for each occupation group, to which the results are sensitive. For example, as evidenced from a comparison of Figure 3.4 and Appendix 3.3 year-on-year there are only small differences in the age-earnings profiles of each occupation group and these lead to the difference in results across years. The upshot is that the results for the individual years are less important that the general finding across all years that nurse training yields the highest NPV if the individual's MTPR is 8%-12% or more. This general finding is explained by the greater disposable income earned by nurses from age 18 to 30, as shown by the age-earnings profiles in Figure 3.4 and Appendix 3.3. While nurses' disposable income falls relative to teachers and graduates after this time, the impact of this is limited with a higher MTPR which has the effect of discounting heavily the relative gains of teachers and graduates over nurses at older ages. It should also be noted that at higher MTPRs not investing might be the preferred option. This occurs if at an MTPR the NPV of all three investments is negative relative to the opportunity cost earnings. The implication is that none of the three investments should be undertaken. In 1996 taking non-manual workers as the opportunity cost investment this occurs at an MTPR of 35%-40% (and probably higher). This outcome is explained by examination of the age-earnings profiles in Figure 3.4 and Appendix 3.3. At younger ages while the investment in human capital is being undertaken the wage benefits of the opportunity cost option outweigh those of the human capital investment.

	Opportunity cost earnings = non-manual workers [1]					
MTPR	1991	1992	1993	1994	1995	1996
0	Graduates	Teachers	Teachers	Teachers	Teachers	Teachers
1	Graduates	Teachers	Graduates	Teachers	Teachers	Teachers
2	Graduates	Teachers	Graduates	Graduates	Graduates	Graduates
3	Graduates	Teachers	Graduates	Graduates	Graduates	Graduates
4	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
5	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
6	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
7	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
8	Graduates	Graduates	Graduates	Graduates	Graduates	Nurses
9	Graduates	Graduates	Graduates	Graduates	Graduates	Nurses
10	Graduates	Graduates	Graduates	Graduates	Nurses	Nurses
11	Graduates	Graduates	Graduates	Nurses	Nurses	Nurses
12	Nurses	Nurses	Graduates	Nurses	Nurses	Nurses
13	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
14	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
15	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
16	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
17	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
18	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
19	No investment	Nurses	Nurses	Nurses	Nurses	Nurses
20	No investment	Nurses	Nurses	Nurses	Nurses	Nurses
21	No investment	No investment	Nurses	Nurses	Nurses	Nurses
22	No investment	No investment	No investment	Nurses	Nurses	Nurses
23	No investment	No investment	No investment	Nurses	Nurses	Nurses
24	No investment	No investment	No investment	Nurses	Nurses	Nurses
25	No investment	No investment	No investment	Nurses	Nurses	Nurses
26	No investment	No investment	No investment	Nurses	Nurses	Nurses
27	No investment	No investment	No investment	Nurses	Nurses	Nurses
28	No investment	No investment	No investment	Nurses	Nurses	Nurses
29	No investment	No investment	No investment	No investment	Nurses	Nurses
30	No investment	No investment	No investment	No investment	Nurses	Nurses
31	No investment	No investment	No investment	No investment	Nurses	Nurses
32	No investment	No investment	No investment	No investment	Nurses	Nurses
33	No investment	No investment	No investment	No investment	Nurses	Nurses
34	No investment	No investment	No investment	No investment	Nurses	No investment
35	No investment	No investment	No investment	No investment	Nurses	No investment
36	No investment	No investment	No investment	No investment	Nurses	No investment
37	No investment	No investment	No investment	No investment	Nurses	No investment
38	No investment	No investment	No investment	No investment	Nurses	No investment
39	No investment	No investment	No investment	No investment	No investment	No investment
40	No investment	No investment	No investment	No investment	No investment	No investment

[1] Where 'No investment' is stated this means that at that MTPR the NPV of all three investments is negative relative to the opportunity cost earnings (in this case non-manual workers). The implication is that none of the three investments should be undertaken.

Table 3.2. Investment with highest NPV, by MTPR, where the opportunity cost earnings are those of non-manual workers

	Opportunity cost earnings = all workers [1]					
MTPR	1991	1992	1993	1994	1995	1996
0	Graduates	Teachers	Teachers	Teachers	Teachers	Teachers
1	Graduates	Teachers	Graduates	Teachers	Teachers	Teachers
2	Graduates	Teachers	Graduates	Graduates	Graduates	Graduates
3	Graduates	Teachers	Graduates	Graduates	Graduates	Graduates
4	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
5	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
6	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
7	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
8	Graduates	Graduates	Graduates	Graduates	Graduates	Nurses
9	Graduates	Graduates	Graduates	Graduates	Graduates	Nurses
10	Graduates	Graduates	Graduates	Graduates	Nurses	Nurses
11	Graduates	Graduates	Graduates	Nurses	Nurses	Nurses
12	Nurses	Nurses	Graduates	Nurses	Nurses	Nurses
13	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
14	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
15	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
16	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
17	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
18	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
19	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
20	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
21	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
22	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
23	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
24	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
25	No investment	Nurses	Nurses	Nurses	Nurses	Nurses
26	No investment	Nurses	Nurses	Nurses	Nurses	Nurses
27	No investment	Nurses	Nurses	Nurses	Nurses	Nurses
28	No investment	No investment	Nurses	Nurses	Nurses	Nurses
29	No investment	No investment	Nurses	Nurses	Nurses	Nurses
30	No investment	No investment	No investment	Nurses	Nurses	Nurses
31	No investment	No investment	No investment	Nurses	Nurses	Nurses
32	No investment	No investment	No investment	Nurses	Nurses	Nurses
33	No investment	No investment	No investment	Nurses	Nurses	Nurses
34	No investment	No investment	No investment	Nurses	Nurses	Nurses
35	No investment	No investment	No investment	Nurses	Nurses	Nurses
36	No investment	No investment	No investment	No investment	Nurses	Nurses
37	No investment	No investment	No investment	No investment	Nurses	Nurses
38	No investment	No investment	No investment	No investment	Nurses	Nurses
39	No investment	No investment	No investment	No investment	Nurses	Nurses
40	No investment	No investment	No investment	No investment	Nurses	Nurses

[1] Where 'No investment' is stated this means that at that MTPR the NPV of all three investments is negative relative to the opportunity cost earnings (in this case all workers). The implication is that none of the three investments should be undertaken.

Table 3.3. Investment with the highest NPV, by MTPR, where the opportunity cost earnings are those of all workers

	Opportunity cost earnings = workers whose highest qualification is A level [1]					
MTPR	1991	1992	1993	1994	1995	1996
0	Graduates	Teachers	Teachers	Teachers	Teachers	Teachers
1	Graduates	Teachers	Graduates	Teachers	Teachers	Teachers
2	Graduates	Teachers	Graduates	Graduates	Graduates	Graduates
3	Graduates	Teachers	Graduates	Graduates	Graduates	Graduates
4	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
5	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
6	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
7	Graduates	Graduates	Graduates	Graduates	Graduates	Graduates
8	Graduates	Graduates	Graduates	Graduates	Graduates	Nurses
9	Graduates	Graduates	Graduates	Graduates	Graduates	Nurses
10	Graduates	Graduates	Graduates	Graduates	Nurses	Nurses
11	Graduates	Graduates	Graduates	Nurses	Nurses	Nurses
12	Nurses	Nurses	Graduates	Nurses	Nurses	Nurses
13	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
14	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
15	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
16	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
17	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
18	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
19	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
20	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
21	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
22	Nurses	Nurses	Nurses	Nurses	Nurses	Nurses
23	Nurses	Nurses	No investment	Nurses	Nurses	Nurses
24	Nurses	No investment	No investment	Nurses	Nurses	Nurses
25	Nurses	No investment	No investment	Nurses	Nurses	Nurses
26	Nurses	No investment	No investment	Nurses	Nurses	Nurses
27	Nurses	No investment	No investment	Nurses	Nurses	Nurses
28	No investment	No investment	No investment	Nurses	Nurses	Nurses
29	No investment	No investment	No investment	Nurses	Nurses	Nurses
30	No investment	No investment	No investment	Nurses	Nurses	Nurses
31	No investment	No investment	No investment	Nurses	No investment	Nurses
32	No investment	No investment	No investment	Nurses	No investment	Nurses
33	No investment	No investment	No investment	Nurses	No investment	Nurses
34	No investment	No investment	No investment	Nurses	No investment	Nurses
35	No investment	No investment	No investment	Nurses	No investment	Nurses
36	No investment	No investment	No investment	Nurses	No investment	Nurses
37	No investment	No investment	No investment	Nurses	No investment	Nurses
38	No investment	No investment	No investment	Nurses	No investment	Nurses
39	No investment	No investment	No investment	Nurses	No investment	Nurses
40	No investment	No investment	No investment	Nurses	No investment	Nurses

[1] Where 'No investment' is stated this means that at that MTPR the NPV of all three investments is negative relative to the opportunity cost earnings (in this case workers whose highest qualification is A level). The implication is that none of the three investments should be undertaken.

Table 3.4. Investment with the highest NPV, by MTPR, where the opportunity cost earnings are those of workers whose highest qualification is A level

For example, across ages 18 to 21 earnings of workers are greater than the bursary for trainee nurses. At older ages when nurses complete their training their earnings become greater. However, the magnitude of this superiority is diminished at higher MTPRs via the effects of the discounting process.

The full sets of private NPV calculations on which these results are based are presented in Appendix 3.4. These show the private NPV of the three different investments relative to the different opportunity cost options for each year, 1991 to 1996. The appropriate decision rule for the individual is to select the investment with the highest NPV at his or her own particular MTPR. When the investment with the highest NPV has in fact a negative NPV then the appropriate decision for the individual is not to invest – the individuals is better off choosing the opportunity cost option.

The private IRRs to nurse training, teacher training and obtaining a degree are presented in Table 3.4. Taking as the baseline opportunity cost earnings represented by female non-manual workers, the private IRR to nurse training was 18%-38% across the period. The private IRR to nurse training was greater than that for teachers (9%-13%) and graduates (13%-15%) in every year (in each case the ranking in descending order by private IRR was nurses then graduates then teachers). This ranking was the same when earnings of non-manual workers and females whose highest academic qualifications are A levels are used as the opportunity cost earnings.

It is not universally the case that the private IRR is lowest using earnings of female non-manual workers as opportunity cost earnings, and highest using earnings of females whose highest academic qualifications were A levels. One might expect that this would be the case given that on average across workers of all ages the mean earnings of non-manual workers are greater than the mean earnings of all workers which in turn are greater than the mean earnings of workers whose highest qualification is A level (see Figure 3.3). Because the private IRR is not lowest when the comparison group with the highest mean earnings is used (non-manual workers) this means that across the entire age-earnings profile at every age it is not always the case that non-manual workers earn more than all workers combined who in turn earn more than workers whose highest qualification is A level. This is borne out by inspection of Figure 3.4 and Appendix 3.3.

	1991	1992	1993	1994	1995	1996
Opportunity cost earnings = non-manual workers						
Nurses	18.4	20.9	21.5	28.3	38.0	33.7
Teachers	9.5	11.6	11.7	12.5	11.9	10.9
Graduates	14.1	14.9	14.7	13.2	14.7	12.5
Opportunity cost earnings = all workers						
Nurses	24.5	27.5	29.8	36.0	43.6	40.9
Teachers	11.7	13.7	13.9	14.7	14.0	13.2
Graduates	16.4	17.3	17.0	15.2	16.9	14.8
Opportunity cost earnings = workers whose highest qualification is A level						
Nurses	27.8	23.2	22.5	40.3	30.5	43.8
Teachers	12.8	13.3	13.3	14.2	11.9	15.1
Graduates	17.6	16.5	15.9	14.9	14.4	16.8

Table 3.5. The private IRR to becoming a nurse and teacher and obtaining a degree in Great Britain, 1991-1996

It should be borne in mind that, as explained above and in greater detail in Appendix 3.2, it is not necessarily the case that the option with the greatest private IRR (in this case, nursing) should be preferred. The usefulness of the private IRR in human capital investment decisions is limited when a comparison of mutually exclusive investments is required and there is a crossover MTPR such as r_C in Figure A.3.2 in Appendix 3.1. Figure 3.5 plots the NPV of nurse training, teacher training and obtaining a degree with non-manual workers as opportunity cost earnings by MTPR in 1996. Trends in NPV by MTPR for 1991 to 1995 with non-manual workers as opportunity cost earnings (see Appendix 3.5) and for all years with the alternative two sets of opportunity cost earnings are similar. The IRR is given by the MTPR where the NPV equals zero (34% for nurse training, 11% for teacher training and 13% for obtaining a degree). However, while the IRR for nurse training is the highest, the NPV for nurse training is highest only when the MTPR is greater than crossover rate (in this case nurse training has a higher NPV than both teacher training and obtaining a degree when the MTPR is 8% of greater). This illustrates the problems arising from a simple comparison of IRRs when a ranking is required of different possible investments. The upshot is that the NPV and the IRR do potentially lead to inconsistent results concerning the relative attractiveness of investments in human capital. In this case the NPV results should be preferred.

The results of the sensitivity analysis on hours worked are presented in Tables 3.6 and 3.7. In terms of the NPV results (Table 3.6) the effect is dramatic because, broadly speaking, teaching becomes the most attractive option at all MTPRs where an investment should be undertaken. Only at very specific MTPRs in some of the years considered does nursing or obtaining a degree become the most preferred option. The rationale behind this substantial change to the results is clear. From Table 3.1 above we can see that relative to the other occupation groups teachers have low recorded mean hours of work per week (only 29 to 30 hours per week). This means that the upward adjustment to teachers earnings that occurs when multiplying mean earnings by a factor 40/h is greatest for teachers. Hence the adjustment has the greatest positive effect on teachers relative mean earnings, and as a consequence of this at all MTPRs the NPV to teaching improves relative to nursing and obtaining a degree. However, these findings should be treated with caution. As explained above the main problem in making adjustments of this kind to the basic earnings profiles is in ensuring comparability of the hours of work data. The hours of work figures for teachers reflect the number of contracted hours that a typical teacher is expected to be in attendance at school. This is likely to underestimate significantly the actual hours of work they undertake each week. For example, hours worked after school hours, organising out-of-school or after-school activities, meeting with parents, preparation time, marking time and time spent writing reports are not included. A more comprehensive analysis taking these factors into account would probably show that teachers on average work considerably longer hours in total than the data used in the sensitivity analysis suggest. In this case the impact on the results would not be as great as shown in Table 3.6.

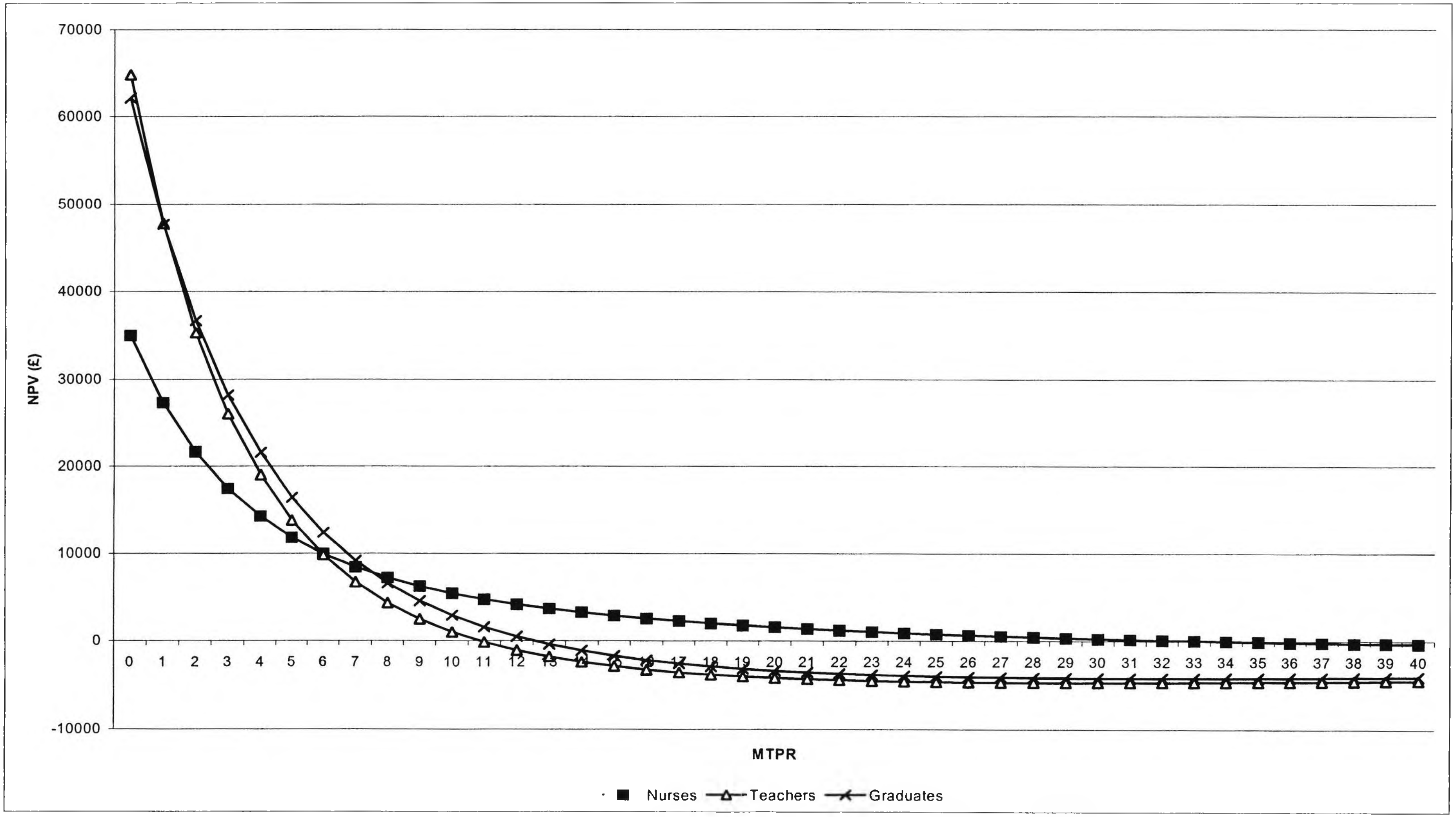


Figure 3.5. NPV of nurse training, teacher training and obtaining a degree with non-manual workers as opportunity cost earnings by MTPR in 1996

	Opportunity cost earnings = non-manual workers [1]					
MTPR	1991	1992	1993	1994	1995	1996
0	Teachers	Teachers	Teachers	Teachers	Teachers	Teachers
1	Teachers	Teachers	Teachers	Teachers	Teachers	Teachers
2	Teachers	Teachers	Teachers	Teachers	Teachers	Teachers
3	Teachers	Teachers	Teachers	Teachers	Teachers	Teachers
4	Teachers	Teachers	Teachers	Teachers	Teachers	Teachers
5	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
6	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
7	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
8	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
9	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
10	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
11	Teachers	Graduates	Teachers	Teachers	Teachers	Teachers
12	Teachers	Nurses	Teachers	Teachers	Teachers	Teachers
13	Teachers	Nurses	Teachers	Teachers	Teachers	Teachers
14	Teachers	No investment	Teachers	Teachers	Teachers	Teachers
15	Teachers	No investment	Teachers	Teachers	Teachers	Teachers
16	Teachers	No investment	Teachers	Teachers	Teachers	Teachers
17	Teachers	No investment	Teachers	Teachers	Teachers	Teachers
18	Teachers	No investment	Teachers	Teachers	Teachers	Teachers
19	No investment	No investment	Teachers	Teachers	Teachers	Nurses
20	No investment	No investment	Teachers	Teachers	Nurses	Nurses
21	No investment	No investment	No investment	Teachers	Nurses	Nurses
22	No investment	No investment	No investment	No investment	Nurses	Nurses
23	No investment	No investment	No investment	No investment	Nurses	Nurses
24	No investment	No investment	No investment	No investment	Nurses	No investment
25	No investment	No investment	No investment	No investment	No investment	No investment
26	No investment	No investment	No investment	No investment	No investment	No investment
27	No investment	No investment	No investment	No investment	No investment	No investment
28	No investment	No investment	No investment	No investment	No investment	No investment
29	No investment	No investment	No investment	No investment	No investment	No investment
30	No investment	No investment	No investment	No investment	No investment	No investment
31	No investment	No investment	No investment	No investment	No investment	No investment
32	No investment	No investment	No investment	No investment	No investment	No investment
33	No investment	No investment	No investment	No investment	No investment	No investment
34	No investment	No investment	No investment	No investment	No investment	No investment
35	No investment	No investment	No investment	No investment	No investment	No investment
36	No investment	No investment	No investment	No investment	No investment	No investment
37	No investment	No investment	No investment	No investment	No investment	No investment
38	No investment	No investment	No investment	No investment	No investment	No investment
39	No investment	No investment	No investment	No investment	No investment	No investment
40	No investment	No investment	No investment	No investment	No investment	No investment

[1] Where 'No investment' is stated this means that at that MTPR the NPV of all three investments is negative relative to the opportunity cost earnings (in this case non-manual workers). The implication is that none of the three investments should be undertaken.

Table 3.6. Investment with highest NPV, by MTPR, where the opportunity cost earnings are those of non-manual workers, standardised to 40 working hours per week

	1991	1992	1993	1994	1995	1996
Opportunity cost earnings = Non-manual workers						
Nurses	12.6	13.4	13.0	17.8	24.9	23.6
Teachers	18.4	10.0	20.6	21.1	20.7	19.3
Graduates	10.4	12.3	11.3	9.1	10.6	7.4

Table 3.7. The private IRR to becoming a nurse and teacher and obtaining a degree in Great Britain, 1991-1996, standardised to 40 working hours per week

3.5. Conclusion

In this chapter we estimate the private NPV and the private IRR to becoming a nurse in Great Britain for the period 1991-1996. Using the NPV investment criterion, nursing is the preferred option for individuals with an MTPR of 8%-12% or more. This range is high relative to the market interest rate and most discount rates recommended for use in public sector investments. Two important questions arising from this analysis are 'what are the factors affecting an individual's MTPR?' and 'which individuals might have an MTPR of 8%-12% or more?' Bearing in mind the caveats raised in Appendices 3.1 and 3.2 concerning the theoretical and practical difficulties, respectively, in estimating the NPV and IRR one possible response to these questions stem from the well-known reasons for incorporating discounting in project appraisal (see, for example, Goodin, 1976). The principle of risk and uncertainty means that individuals will apply a higher MTPR if their financial future or their family's financial future is more uncertain. The principle of the diminishing marginal utility of income, which implies that the marginal utility of money will increase as less is held, implies that the less well-off the person, or the poorer their family background, the higher the MTPR they are likely to apply. The principle of the opportunity cost of investment means that an individual is likely to apply a higher MTPR the higher the opportunity cost of the investment, which is likely to be the case for individuals from poorer backgrounds with less investment capital. The above factors all point to the MTPR being higher for individuals from poorer backgrounds. One might then hypothesise that, for example, such individuals would find nursing a career an attractive option.

We also estimate the private IRRs to nurse training, and find these to be greater than those for teacher training and obtaining a degree. However, we also show that ranking investments using the IRR criterion is inappropriate when there exists a crossover MTPR, which is shown to be the case here. In this instance the NPV is superior to the IRR in ranking human capital investments.

There are a number of issues that may affect the inferences to be drawn from the results. First, the methods used in this analysis are data dependent. We would like to have considered different careers within the nursing profession, other comparator occupations and a longer time period, but a lack of reliable data prevented us from doing this. Also due to data constraints we have calculated the private NPV and private IRR for females only. In the context of the British nursing labour market this is not particularly problematic since a high proportion of nurses (some 90%) are females.

Second, if it is the case that individuals select occupations in which they have a comparative advantage in terms of natural ability and motivation then comparing mean earnings of nurses by age to those in other occupations in a model of occupational choice may be unrealistic. Difficulties arise in choosing the appropriate level of opportunity cost earnings from undertaking training. In this analysis we calculate the NPV and IRR using a range of opportunity cost profiles. Difficulties also arise comparing different investments in human capital. For example, it might not be the case that a nurse would earn the average wage of a teacher if that nurse became a teacher. This comparative advantage effect means that average earnings data may not accurately represent the benefits to an individual of undertaking an alternative investment in human capital.

Third, computation of age-earnings profiles ideally requires data on future earnings at older ages. These are unobservable. Hence, the method adopted in this and other studies is to use cross-sectional data on mean earnings of individuals at each age and in each occupation/qualification group, assuming that current earnings of individuals at older ages will reflect future earnings of individuals at older ages. This may lead to conflation of age and cohort effects and we present some evidence of this using 1980 earnings data. Unfortunately, precisely because future earnings are unobservable it is difficult to estimate the true magnitude of this conflation for the time periods considered in the analysis (1991 to 1996).

In summary, in Chapter 3 we have estimated the private net present value and private internal rate of return to becoming a nurse in Great Britain. The calculations are made using the standard equations inputted with data from the NES and the BHPS. Basic age-earnings profiles are adjusted for mortality, unemployment, other causes of economic inactivity, and discontinuation from training. The conclusions are that: (1) there is a high private internal rate of return to becoming a nurse in Great Britain relative to other occupations; (2) using the internal rate of return criterion is inappropriate when there exists a crossover marginal time preference rate, which is shown to be the case here; and, (3) using the net present value criterion there are net financial benefits to becoming a nurse in Great Britain for individuals with a marginal time preference rate of 9%-14% or more.

The implication of the analysis is that on financial grounds in terms of their earnings there is a rationale for choosing to be employed as a nurse in Great Britain. While in terms of relative earnings a career in nursing may not have universal appeal (depending on the MTPR) for a section of the population with a relatively high discount rate nursing is an attractive option in this regard.

CHAPTER 4

A THEORETICAL MODEL OF EARNINGS FOR NURSES AND A REVIEW OF THE LITERATURE

4.1. Introduction

In Chapters 4 to 6 we build on the analysis of the previous chapters and examine the factors that affect the wages of nurses working in the NHS. We also examine the nature and magnitude of wage differentials between nurses and other workers and investigate the causes of these observed differentials. Chapter 4 consists of two main sections. In the first we outline the economic and statistical models to be estimated in Chapters 5 and 6. In the second main section we review the literature to date on earnings function for nurses.

We begin the chapter by providing an exposition of earnings functions as an empirical tool for the analysis of the determinants of wages. Generically, the term ‘earnings function’ has come to mean any regression of individual wage rates or earnings on a vector of personal, market and environmental variables thought to influence wages (Willis, 1986). The main application of earnings functions is to examine the effects of investments in human capital (particularly schooling and the attainment of educational qualifications) on earnings. Jacob Mincer (1974) conducted the pioneering work in this area. This chapter is devoted to the theoretical and empirical development of the human capital earnings function stemming from the work of Mincer. We construct a model of earnings that will form the basis for the empirical estimation of an earnings function for nurses in Chapters 5 and 6. We begin by examining the Mincerian earnings function. We then introduce a key estimation issue in applying the basic Mincerian model to the labour market for nurses in the NHS, namely the problem of selection bias. Having described and explained the significance of this problem we move on to discuss potential solutions, based primarily on the work of James Heckman (Heckman, 1979). We finally survey empirical applications of earnings functions applied to the labour market for nurses.

4.2. The Mincerian earnings function

The standard human capital earnings function developed by Mincer (1974) is of the form:

$$\ln Y = \beta_0 + \beta_1 S + \beta_2 t + \beta_3 t^2 \qquad [4.1]$$

The derivation of this function is given in Appendix 4.1. lnY is the natural logarithm of actual observed earnings, S represent years of schooling, and t represents years of post-school work experience. β_1, the co-efficient on schooling, provides an estimate of the (constant) rate of return to education. A concavity in the earnings profile with respect to years of post-school work experience is captured by the quadratic experience terms t and t^2. One would expect the co-efficients on these terms (β_2 and β_3, respectively) to be positive and negative, respectively, to capture an n-shaped experience-earnings profile.

Equation [4.1] was fitted by Mincer to data by regression analysis. We shall now look at the results obtained by Mincer and interpret the co-efficients. This is useful as an empirical exercise but also because the results are important in understanding the earnings process with regards to human capital accumulation. Table 4.1 contains earnings functions estimated using 1960 US Census data for white non-farm-working males. In the data S was measured directly but because the Census data did not record workers' actual labour force experience a transformation of the worker's age A was used instead. Mincer used the transformation t = A – S – 6, which assumes that a worker begins full-time work immediately after completing his education and that the age of school completion is S + 6.

Equation form and results [1-2]	R^2
1. $\ln Y = 7.58 + 0.070S$	0.067
2. $\ln Y = 6.20 + 0.107S + 0.081t - 0.0012t^2$	0.285
3. $\ln Y = 4.87 + 0.255S - 0.0029S^2 - 0.0043tS + 0.148t - 0.0018t^2$	0.309

[1] All co-efficients statistically significant at 5% level
[2] Y = annual earnings of white US males in 1959 US$, S = years of schooling completed, t = age – years of schooling completed – 6
Source: Mincer (1974)

Table 4.1. Estimates of human capital earnings functions

Estimates by Mincer of three specifications of the human capital earnings function are presented. Line 1 shows an estimate of a 'schooling model' that assumes no post-school investments in human capital. The estimated rate of return to one extra year of schooling is 7%. This equation explains only 6.7% of the variance in ln annual earnings.

Omitting experience from the earnings function as in line 1 results in a downward bias in the co-efficient on schooling because schooling and experience tend to be negatively correlated due to the fact that at any given age those with more schooling are likely to have less experience. Suppose we compare two individuals of the same age (say, 28 years). Let one individual have 15 years of schooling (from ages 6 to 21) and the other 12 (from ages 6 to 18). The 15-year individual would then have at most 7 years of experience, while the other individual would have at most 10 years. Both schooling and experience increase earnings. In comparing these two individuals one has more schooling but less experience and the other has more experience and less schooling. Fitting an earnings function such as that in line 1 which relates earnings only to schooling neglects the fact that the more educated have on average less experience. Therefore, this earnings function mistakenly neglects to account for differences in experience for individuals who are of the same age but have different schooling levels. Moreover, neglecting experience implies that this factor has no effect on earnings so that the experience-earnings profile is flat. Consequently the intercept in line 1 represents average earnings across all experience groups.

The extent of the downward bias from omitting experience from the regression is illustrated in line 2, which presents an earnings function of the form given by equation [4.1]. In this specification the estimated rate of return to one extra year of schooling rises to 10.7%. The co-efficients on experience and experience squared imply that the experience-earnings profile is n-shaped. The slope of this profile is $\frac{\delta \ln Y_t}{\delta t} = 0.081 - 0.0024t$. For $t < 33.75$ this is positive. The second derivative is –0.0024, which indicates that earnings rise with experience but at a diminishing rate, reaching a maximum at an experience level of 33.75 years. The intercept term (6.20) depicts ln annual earnings for an individual with no schooling. The antilog of 6.20 is 493. In part this low figure is because the data relate to 1959 US$. In part it is also due to the fact that we are extrapolating outside of the sample space – that is, very few individuals have no schooling whatsoever. Suppose we were to take, using Mincer's assumptions, the average schooling level to be 12 years (from age 6 to 18 years). Then one could expect to augment ln annual earnings by 0.107*12 = 1.284 by including schooling effects. ln annual earnings would then be 7.484, which translates into earnings of $1,772 in 1959 prices. Note that the addition of the experience terms increases markedly the explanatory power of the regression, raising R^2 to 28.5%.

The earnings function in line 3 has two additions to the basic earnings function presented in line 2. First, schooling squared is added to allow for non-linearities in the rate of return to schooling. In this case the positive and negative co-efficients on S and S^2, respectively, indicate that the schooling-earnings profile is n-shaped (a diminishing rate of return with increased schooling – just as with experience). In this case the return to an extra year of schooling is $\frac{\delta \ln Y_t}{\delta S} = 0.255 - 0.0058S - 0.0043t$. This implies a diminishing rate of return to years of schooling with increased schooling as well as with increasing work experience. For an individual with no experience and 10 years of schooling the rate of return to an extra year of schooling is 0.255 – 0.0058(10) – 0.0043(0) = 0.197 or 19.7%. Suppose now an individual with no experience and 15 years of schooling. In this case the rate of return to an extra year of schooling is 0.255 – 0.0058(15) – 0.0043(0) = 0.163 or 16.3%. Therefore, line 3 indicates a declining marginal rate of return to years of schooling as years of schooling rise. This specification also shows that the marginal rate of return to schooling diminishes for individuals with more labour market experience.

The interaction term (tS) in line 3 also allows the experience-earnings profile to vary with schooling. In line 2 the rate of which earnings rise diminished solely with experience, now the slope is also related to the schooling level. $\frac{\delta \ln Y_t}{\delta t} = 0.148 - 0.0036t - 0.0043S$. This means that experience-earnings profiles increase more slowly the greater the level of schooling.

Earnings functions like those reported in Table 4.1, especially those of the form given in line 2 (i.e. equation [4.1]), have been estimated hundreds of times using both cross-sectional and longitudinal data sources for many countries (Willis, 1986). The important point is that the Mincerian earnings function has been proved many times to be consistently useful as an empirical tool for the analysis of the determinants of wages. As noted by Willis (1986): "As an empirical tool the Mincer earnings function has been one of the great success stories of modern labour economics. It has been used in hundreds of studies using data from virtually every historical period and country for which suitable data exist. The results of these studies reveal important empirical regularities in educational wage differentials and the life cycle pattern of earnings."

4.3. Extended earnings functions

Various modifications to the basic earnings function constructed by Mincer are possible. The most common are the inclusion of exogenous variables known to, likely to or that will possibly affect earnings (see, for example, Polachek and Siebert, 1993, for a review). The following is a (by no means exhaustive) list of variables that have been used in the construction of earnings functions in addition to years of schooling and years of experience:

- Sex;
- Marital status;
- Ethnic group;
- Nationality;
- Hours worked per week;
- Weeks worked per year;
- Region of residence;
- City/population size;
- Level of health/disability;
- Age;
- Member of a union;
- Industry;

- Occupation;
- Number of children;
- Number of jobs since left school;
- Starting salary;
- Type of school attended;
- Education attainment/qualifications obtained;
- Social class;
- Religion;
- Family wealth; and
- Mother's and father's education.

In addition, interaction terms between these variables as well as the others already studied above are possible. The inclusion of variables such as these to the basic Mincerian earnings function has allowed the model to be expanded in an impressive array of directions to address a number of different questions. Some examples include:

1. What are the magnitude and causes of earnings differentials across ethnic groups in the labour market? (Hamilton, 1997, Jones, Nadeau and Walsh, 1999, Cabezas and MacDonald, 1999, Gyimah-Brempong and Fichtenbaum, 1999.)
2. What are the magnitude and causes of male-female earnings differentials in the labour market? (Dolton and Makepeace, 1986, Schuld, Schippers and Siegers, 1994, Joshi and Paci, 1998, Sicilian and Grossberg, 2001.
3. Are individuals who work in the private sector paid more than those who work in the public sector? (Kanellopoulos, 1997, Prescott and Wandschneider, 1999.)

4. Do immigrant workers earn less than native-born workers? (Shields MA. And Wheatley Price, 1998, Cabezas V. and MacDonald, 1999.)
5. What are the rates of return to on-the-job training? (Groot and Mekkeholt, 1995.)
6. Do past earnings affect current earnings capabilities? (Parker, 1994, Cabezas and MacDonald, 1999.)
7. What is the hourly wage differential between full-time and part-time workers? (Harris, 1993.)
8. What are the returns to being overeducated in an occupation? (Groot, 1996.)
9. What is the relationship between religious denomination and earnings? (Tomes, 1985.)
10. Do wages compensate for the job risk or individual risk of unemployment? (Moretti, 2000.)
11. How sensitive are earnings functions to different measures of work experience? (Lambert, 1993.)
12. What characteristics of higher education institutions affect earnings? (James et al., 1989.)
13. How is expected employment tenure related to earnings? (Tanaka, 2001.)
14. What are the returns to different levels of education? (Sanmartin, 2001, Palme and Wright, 1998, Williams and Gordon, 1981.)
15. What factors affect the remuneration of workers in specific occupations? (Baimbridge and Simpson, 1996, Hamilton, 1997, Jones, Nadeau and Walsh 1999.)
16. What effect does volunteer work have on earnings? (Day and Devlin, 1997.)
17. What effect does uncertainty in terms of educational requirements to enter an occupation have on earnings? (Robst J. and Cuson-Graham, 1999.)
18. What is the impact of mental job stress on earnings? (French and Dunlap, 1998.)
19. How does the expectation of earnings affect choice of career path? (Johnes, 1999.)

20. How does education and training increase employability and how does this affect wages? (Groot and Van Den Brink, 2000.)
21. Does household labour impact on labour market earnings? (McLennan, 2000.)
22. Do wages become more dispersed over time and what are the causes and consequences of this? (Pagan JA. and Tijerina-Guajardo, 2000.)
23. What effect does temporary withdrawal from the workforce have on earnings? (Simpson, 2000.)

This brief survey of empirical research based on extended human capital earnings functions has only skimmed the surface of a massive literature that utilises this approach to study a wide variety of subjects that are too numerous to be considered in more detail here. See, for example, Willis (1986), Elliott (1991), Killingsworth (1983) and Polachek and Siebert (1993) for further in-depth reviews.

As noted above the premier application of earnings functions is to examine the returns to investments in human capital where additional variables are included in the basic earnings function to capture the effects of the attainment of education (such as years of schooling or obtaining a degree). As pointed out by Elliott (1991) however, there are a number of difficulties that researchers encounter in estimating these returns.

4.4. Difficulties in using earnings functions to estimate rates of return on investments in human capital

4.4.1. On-the-job training

The first problem that researchers encounter in attempting to derive an accurate measure of the rate of return to an investment in human capital is that such investment often continues over much of an individual's working life. The problem is that we may overstate the returns on an initial human capital investment (such as schooling or obtaining a degree or undergoing initial training for a specific occupation such as nursing) if we fail to allow for and distinguish separately the returns to subsequent on-the-job investment. In those occupations where on-the-job training is of a formal nature this can be allowed for by the inclusion of further variables in the earnings function (e.g. 'years of on-the-job training'). However, where such training is of a less formal nature and therefore comprises what has been called 'learning by doing' it is more difficult to allow for its effects.

One obvious and straightforward approach to overcoming this problem is to assume that on-the-job training is related to (and measured by) years of experience in the labour market. This is the approach that is adopted in the present analysis, and is clearly consistent with the Mincerian earnings function presented in equation [4.1]. However, it should be borne in mind that this might not accurately measure the influence of on-the-job training on earnings. There are jobs that require a relatively trivial level of on-the-job training although the individuals who occupy such jobs do so for a long period of time. Experience in this case is a poor measure for the magnitude of additional skills required to do the job. This is unlikely to be the case for nursing where advances in health care technology mean that continual training over an individual's career is required.

4.4.2. Informal information networks

The acquisition of information about potential employee's productive characteristics requires expenditure by employers. Employers may seek to obtain such information at least cost. Ad hoc information networks (such as 'old boy networks') represent informal channels for transmitting information about individuals' productive characteristics and may offer opportunities for employers to economise on their expenditures in the process of acquiring information about potential employees. If employers who pay high wages use channels such as these there is a financial advantage to individuals who gain access to them. If access to the network and therefore to the occupation is gained by social contacts the estimated return to an investment in human capital will be overstated. This is because it will now also include a return to being a member of the informal information network. Unfortunately, as noted by Elliott (1991), no independent measures of these effects have yet been devised.

4.4.3. Non-wage benefits

Earnings functions estimate the returns on investments in human capital in terms of changes in earnings. This is because earnings are the most easily measurable reward for working and comprise the major part of individuals' returns from work. However, an individuals' total compensation also includes non-wage factors. Such factors might include non-wage advantages and disadvantages associated with training and education programmes and also with different occupations. For example, non-wage effects of education and training might include those experienced while actually undergoing education and training such as enjoying the lifestyle of a student. Non-wage effects arising from working in a specific occupation are likely to be driven by working conditions and job satisfaction, which in turn will be affected by things such as the availability of flexible working hours, the ability to go on training and refresher courses, the provision of child care, working hours, and the availability of career counselling. Failure to include positive non-wage factors means that total returns to human capital investments will be understated, and vice versa. The importance of non-wage factors in nursing will be discussed in more detail in Chapter 7.

4.4.4. Selection bias

The implicit assumption underlying the Mincerian earnings function is that all individuals are equally suited for all jobs. Unfortunately, individuals in reality differ in their ability and motivation in different occupations and part of the differences in observed earnings is attributable to these factors. For example, we might believe that, on average, individuals who are more able, hardworking, energetic and enterprising are more likely to invest in human capital, more likely to participate in the labour market, and also more likely to secure higher paid jobs. Unless we have some way of allowing for the effects of these attributes the estimated return to an investment in human capital may overstate the true return because it will also capture the return to ability. The issue of ability is one aspect of what has come to be known as the selection bias problem, because generally it appears to be the case that those who invest most in human capital are a self-selecting group of the most able.

One solution to the problem is to employ econometric techniques that deal with problems of selection bias. We shall now explore the problem of selection bias and consider the solutions in more depth.

4.5. Selection bias and earnings functions

Suppose we wish to estimate by regression analysis a human capital earnings function of the form specified in equation [4.1]. More appropriately for econometric estimation, let us suppose the equation we wish to estimate is of the form:

$$\ln W_i = \beta X_i + U_i \qquad [4.2]$$

where W is wages, X is a matrix of individual human capital characteristics (including years of schooling S and post-school work experience t, as shown in equation [4.1]) and other exogenous socio-economic variables affecting wages, β is a vector of unknown parameters, U is a normally distributed error term with zero mean and constant variance, σ_U^2.

Ordinary least squares (OLS) is the obvious estimation technique to use to estimate the statistical model described in equation [4.2] based on individual data, but, as noted above, this does not allow for potential selection bias. The traditional selection bias models began with the work of Heckman and others in the 1970s (see Heckman, 1979, Gronau, 1974, and Lewis, 1974). Work in this area has two branches. The first concerns the estimation of statistical models that are observed for only a sub-sample, either by definition – as in the case of wage rates, which are by definition only observed for those who are working – or by fortune of the data available. This type of model has been termed the "partial population" selection model. The second branch presumes that information is observed and available for the total population in the data but that there are one or more regressors of interest that take on their values as a result of some kind of selection process. The simplest case assumes interest to centre on a single dummy variable for some type of 'treatment' – such as being employed in a specific occupation, such as nursing – and hence this type of model is often referred to as the "treatment effect" selection model. See Moffitt (1999) for a further discussion of the distinction between the two types of selection model.

In the context of examining the factors that affect the wages of nurses working in the NHS we focus first on selection bias arising through the decision to participate in the labour market. This is a potentially serious problem because wages rates are by definition observed only for individuals who choose to participate in the labour force. Estimating the earnings function for workers without allowing for potential correlation between the decision to work and the wage will yield biased coefficient estimates if non-workers are not randomly selected from the sample. To see this more clearly we first need to look more closely at the labour supply decisions of individuals – we need to develop a model of individual labour supply.

4.6. A model of individual labour supply

The following is based on Killingsworth (1983). Suppose an individual's utility function is defined as follows:

$$U = U(C, L) \qquad [4.3]$$

where U is utility, C is the amount of market goods consumed, and L is the amount of leisure consumed per time period. The individual aims to maximise his or her utility subject to a budget constraint of the form:

$$WH + N = PC \qquad [4.4]$$

where W is the wage rate per hour, H is the number of hours worked, N is income from sources unrelated to work (property income), and P is the price of a unit of C. To simplify the exposition let us first redefine W and N to be the real hourly wage and the real level of property income, respectively. This means that P may effectively be dropped from the discussion. The total available time per period, T, may be allocated between leisure and work:

$$T = H + L \qquad [4.5]$$

The utility maximisation problem of the individual is therefore defined as follows:

$$\text{Maximise } U = U(C, L) \text{ subject to } WH + N = C \qquad [4.6]$$

Given this problem the determination of labour supply is shown using Figure 4.1.

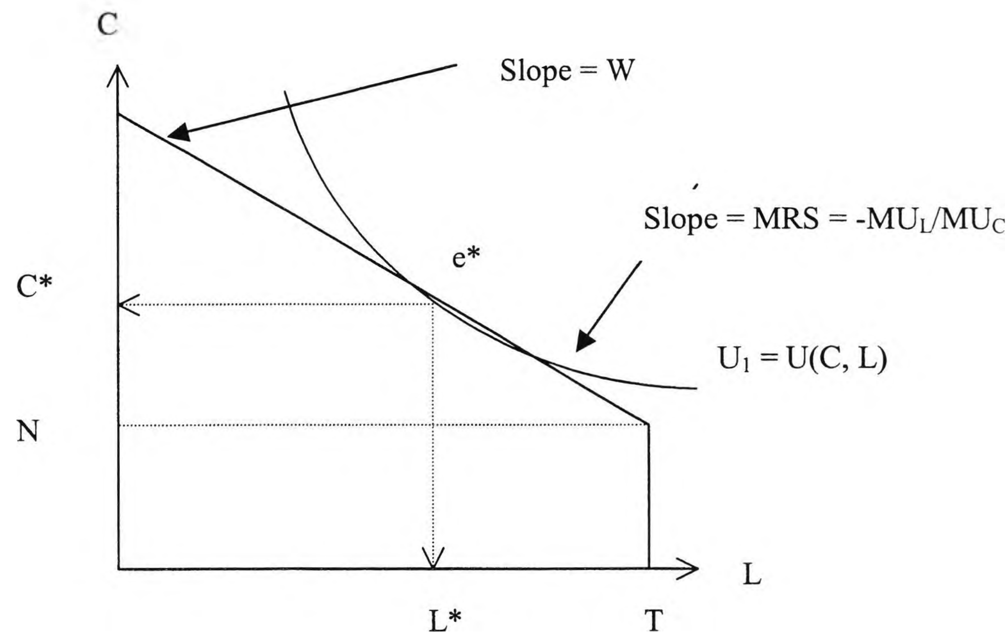


Figure 4.1. Determination of individual labour supply

The indifference curve U_1 shows different combinations of C and L that yield the same level of utility to the individual. The slope of the indifference curve, the marginal rate of substitution (MRS), is given by:

$$MRS = -\frac{MU_L}{MU_C} \quad [4.7]$$

The constraints facing the indifference curve are summarised graphically by the budget line. The individual receives a real property income N and receives a real wage of W for each hour worked. The optimal combination of C and L is the one lying on the highest attainable indifference curve, shown by C*L* at point e*. The equilibrium condition is given by:

$$\frac{MU_L}{MU_C} = W \quad [4.8]$$

The optimum at e* is an interior solution since C, L and H are all positive. At lower values of W or at higher values of N a corner solution (where H = 0) may be optimal. For example, as shown in Figure 4.2, suppose that the real hourly wage falls to W_2, the optimum now becomes point f* where L = T. If the wage fell still further then the slope of the budget line would remain less than the slope of indifference curve U_2 and the corner solution would remain. W_2 is therefore the (real) reservation wage – the highest wage at which the individual will not work.

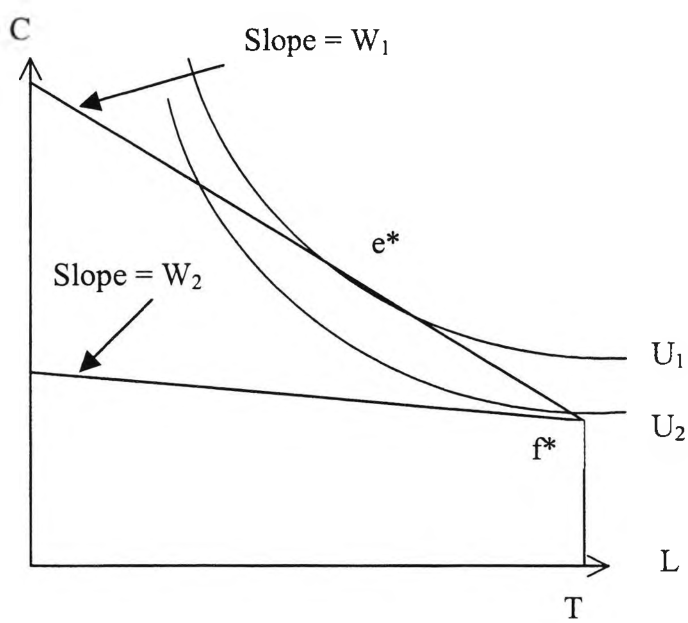


Figure 4.2. Individual labour supply with a corner solution

With this general model in mind suppose we now specify the individual's utility function with a Cobb-Douglas form as follows:

$$U = C^{\alpha}L^{\beta} \qquad [4.9]$$

This is maximized subject to the budget constraint WH + N = C. With this specification of the individual's utility function the MRS and the reservation wage, W_r, are given by:

$$MRS = \left[\frac{b}{(1-b)}\right]\frac{C}{L} = \left[\frac{b}{(1-b)}\right]\frac{(WH+N)}{(1-H)} \qquad [4.10]$$

and

$$W_r = MRS_{L=1} = \left[\frac{b}{(1-b)}\right]N \qquad [4.11]$$

respectively, where $b = \left[\frac{\beta}{(\alpha+\beta)}\right]$. Moreover, the individual will work if the wage he can earn, W, exceeds the value of his reservation wage, W_r:

$$H > 0 \text{ if and only if } W > W_r \qquad [4.12]$$

that is:

$$H > 0 \text{ if and only if } W > \left[\frac{b}{(1-b)}\right]N \qquad [4.13]$$

If the individual does work (and therefore is at a point of indifference curve-budget line tangency, such as point e* in Figure 4.1), then his wage rate W and his marginal rate of substitution MRS are equal at his particular level of hours of work. Because MRS is a function of hours of work H this means that the actual value of labour supply in this case is determined by the condition W = MRS, that is, by solving the relation:

$$W = \left[\frac{b}{(1-b)}\right]\frac{(WH+N)}{(1-H)} \qquad [4.14]$$

for H. This yields the following labour supply function:

$$H = 1 - b - b\left(\frac{N}{W}\right) \qquad [4.15]$$

This simple theoretical model is not suitable for empirical estimation, since it ignores the fact that individuals differ not only in terms of observable variables (N and W) but also in terms of unobservables (a random error term). These unobservables, like observable variables, play a part in labour supply decisions and an appropriate specification of an empirical labour supply function requires not only specification of the role of observable variables but also careful specification of the way in which unobservables affect labour supply decisions.

We can consider a generalisation of the above labour supply model to include an unobservable error term e so as to make the model suitable for application to a population of different individuals. We may now redefine the Cobb-Douglas utility function given by equation [4.9] as follows:

$$U = [W(H + e) + N]^{\alpha}[1 - (H + e)]^{\beta} \qquad [4.16]$$

where e is an unobservable error term that varies from one person to another and that may be interpreted as a “taste shifter” that is, as representing interpersonal differences in tastes for work. Note that equation [4.16] implies that persons with different values of e will not necessarily derive the same utility from given amounts of C (= WH + N) and L (= 1 – H), even if their wage rates and property incomes are the same.

The MRS implied by this utility function is:

$$\mathrm{MRS} = \frac{(\delta U/\delta L)}{(\delta U/\delta C)} = \left[\frac{b}{(1-b)}\right]\frac{[W(H+e)+N]}{[1-(H+e)]} \qquad [4.17]$$

where $b = \left[\frac{\beta}{(\alpha+\beta)}\right]$. W_r is:

$$W_r = \left[\frac{b}{(1-b)}\right]\frac{(eW+N)}{(1-e)} \qquad [4.18]$$

Therefore, an individual with a given value of e will work if and only if $W > W_r$, or:

$$W > \left[\frac{b}{(1-b)}\right]\frac{(eW+N)}{(1-e)} \qquad [4.19]$$

This may be rearranged to:

$$-e = \frac{-(1-b)}{W}\left[W - \frac{b}{1-b}N\right] \qquad [4.20]$$

Or, in other words,

$$H > 0 \text{ if and only if } \varepsilon_H > -\left[(1-b) - b\left(\frac{N}{W}\right)\right] \qquad [4.21]$$

$$H = 0 \text{ if and only if } \varepsilon_H \leq -\left[(1-b) - b\left(\frac{N}{W}\right)\right] \qquad [4.22]$$

where ε_H = -e.

If an individual does work, his hours of work are determined by the condition W = MRS, that is, by solving the relation:

$$W = \left[\frac{b}{(1-b)}\right]\left[\frac{W(H+e)+N}{1-(H+e)}\right] \qquad [4.23]$$

for H. This yields the empirical labour supply function:

$$H = \frac{-(1-b)}{W}\left[W - \frac{b}{1-b}N\right] - e \qquad [4.24]$$

or, equivalently:

$$H = (1 - b) - b\left(\frac{N}{W}\right) + \varepsilon_H \qquad [4.25]$$

Equations [4.21]–[4.25] emphasise an important threshold condition: individual i will work if equation [4.21] is satisfied and will not work if equation [4.22] is satisfied. In behavioural terms, this means that an individual with given N and W will work only if his tastes for work summarised by the magnitude of ε_H are sufficiently greater than the quantity $-\left[(1-b)-b\left(\frac{N}{W}\right)\right]$.

Given this empirical model of labour supply, which includes the threshold condition for participation in the labour market, we can now assess the likely impact of participation selection bias that arises because wage rates are only observed for those who are working.

4.7. Participation selection bias

Individuals choose whether or not to work (to participate in the labour market), and so choose whether or not they earn an observed wage. If individuals make this decision randomly then we could ignore the fact that not all wages are observed and use OLS regression to estimate the earnings function specified by equation [4.2]. Unfortunately such a random participation model is unlikely to be true. Two distinct types of selection bias might occur in this context.

Case 1. Observed wages are biased upwards

An individual i would choose not to work when their personal reservation wage (W_{ri}) is greater than the wage offered by employers (W_i), that is, $W_i < W_{ri}$. Individuals who otherwise would earn relatively low wages (for example due to their human capital characteristics) may be unlikely to choose to work for this reason and therefore the sample of observed wages would be biased upwards.

Case 2. Observed wages are biased downwards

Individuals who choose not to work might have earned even higher wages than those who do choose to work: they may have higher offer wages W_i but an even higher reservation wage W_{ri}. In this case the sample of observed wages is biased downwards.

To examine these issues more formally in a framework suitable for econometric estimation let us assume there are two types of individual: those in paid work and those not in paid work. As we have seen above an individual will decide to undertake paid work if $W_i > W_{ri}$, where:

$$\ln W_{ri} = \beta_r X_{ri} + U_{ri} \qquad [4.26]$$

where X_r is a vector of regressors, β_r is a vector of parameters and U_r is an error term, all pertaining to the reservation wage, W_r. An individual's propensity to participate can therefore be measured by a continuous variable P*, defined as:

$$P_i^* = \ln W_i - \ln W_{ri} = (\beta X_i - \beta_r X_{ri}) + (U_i - U_{ri}) = \delta Z_i + V_i \qquad [4.27]$$

where Z is a vector of regressors, δ is a vector of parameters and V is an error term with zero mean and constant variance, σ_V^2. This is the equation that determines sample selection (participation), and will be referred to as the participation equation. The equation of primary interest is equation [4.2]. The sample selection rule is that $\ln W_i$ is observed only when P_i^* is greater than zero (meaning that wages are observed because the individual chooses to participate only when the offered wage is greater than the reservation wage). Suppose that U_i

and V_i have a bivariate normal distribution with zero means and correlation ρ. The following then applies (based on Heckman, 1979, and Greene, 2000):

$$E(\ln W_i \mid \ln W_i \text{ is observed}) = E(\ln W_i \mid P_i^* > 0) \quad [4.28a]$$

$$= E(\ln W_i \mid V_i > -\delta Z_i) \quad [4.28b]$$

$$= \beta X_i + E(U_i \mid V_i > -\delta Z_i) \quad [4.28c]$$

$$= \beta X_i + \rho\sigma_U \lambda(p)_i \quad [4.28d]$$

$$= \beta X_i + \beta_{\lambda(p)} \lambda(p)_i \quad [4.28e]$$

where

$$\lambda(p)_i = \frac{\phi(-\delta Z_i / \sigma_V)}{1 - \Phi(-\delta Z_i / \sigma_V)} \quad [4.29]$$

where ϕ and Φ are, respectively, the standard normal density function and standard normal cumulative distribution function and the p of $\lambda(p)$ denotes an adjustment for participation selection bias. From equation [4.28e] we therefore have the following:

$$\ln W_i \mid P_i^* > 0) = E(\ln W_i \mid P_i^* > 0) + \varepsilon_i \quad [4.30a]$$

$$= \beta X_i + \beta_{\lambda(p)} \lambda(p)_i + \varepsilon_i \quad [4.30b]$$

Equation [4.30b] shows that OLS regression using the observed data for workers only (equation [4.1]) produces inconsistent estimates of β, and the problem can be viewed as an omitted variable [$\lambda(p)$] problem.

P^* is an unobserved latent variable. In practice we observe only whether an individual is in paid work (denoted by $P = 1$), or not in paid work ($P = 0$). When the value of P^* is positive, the individual decides to undertake paid work, and vice versa. Therefore we may construct a sample selection model as follows:

Selection mechanism: $P_i^* = \delta_i Z_i + V_i$

$P_i = 1$ if $P^*_i > 0$ and 0 otherwise

$\text{Prob}(P_i = 1) = \Phi(\delta Z_i)$ and $\text{Prob}(P_i = 0) = 1\text{-}\Phi(\delta Z_i)$

Regression model: $\ln W_i = \beta X_i + U_i$, observed only if $P_i = 1$,

$(U_i, V_i) \sim$ bivariate normal $(0, 0, 1, \sigma_U, \rho)$

Suppose that P_i and Z_i are observed for a random sample of individuals but that $\ln W_i$ is observed only when $P_i = 1$. This model is precisely the one described above (equation [4.28e]):

$$E(\ln W_i \mid P_i = 1) = \beta X_i + \beta_{\lambda(p)}\lambda(p)_i \qquad [4.31]$$

Estimation of the parameters of the sample selection model described above proceeds in two steps, as described by Heckman (1979):

Step 1: Estimate the probit sample selection equation (equation [4.27]) by maximum likelihood to obtain estimates of δ. Compute $\hat{\lambda}_i$ for each observation in the selected sample. Call this the participation equation.

Step 2: Estimate β and β_λ by OLS regression of $\ln W$ on X and $\hat{\lambda}$. Call this the wage equation.

The resulting adjusted earnings function that is estimated via Step 2 is given by:

$$\ln W_i = \beta X_i + \beta_{\lambda(p)}\lambda(p)_i + \varepsilon_i \qquad [4.32]$$

This method of correcting for participation selection bias is sometimes referred to as the Heckman two-step procedure in recognition of Heckman's (1979) contribution.

The probit model of participation (Step 1) is described in greater detail in Appendix 4.2. This is included to provide an exposition of the nature of the first step of Heckman's two-step model. It also helps us to understand the meaning of the participation selection bias correction term, $\lambda(p)$.

In summary, what the above exposition tells us is that the solution to the participation selection bias problem is to include an additional regressor that corrects the market wage to account for what is essentially the individual's propensity to be in paid work. The correction term is the product of a scalar, $\beta_{\lambda(p)}$, which is constant across all selected cases, and $\lambda(p)$, a term that varies across individuals in the sample. $\beta_{\lambda(p)}$ is a function of the standard deviation of the error term in the earnings function, σ_U, and in the participation equation and their correlation, ρ. $\lambda(p)$ reflects the predicted probability of being in paid work given other known characteristics. It has been referred to as the Inverse Mills Ratio.

$\lambda(p)$ is calculated from a regression of participation status on personal characteristics to generate a participation equation. The dependent variable is a binary discrete variable that takes the value 1 if the individual is in paid work and a value of 0 otherwise. The model can be estimated as a probit model to generate an estimate of $\lambda(p)$. $\lambda(p)$ is then included as an additional regressor in the earnings function. $\beta_{\lambda(p)}$, its coefficient, may have either a positive or negative sign depending on whether the otherwise unmeasured characteristics of those in paid work compared with those not in paid work lead to higher or lower wages (see below for a complete interpretation of λ).

As explained in greater detail in the next chapter, in the empirical analysis the participation equation is estimated using the whole sample of individuals in the data (participators and non-participators), and the wage equation is estimated using the sub-sample of those who participate.

We have now derived an appropriate model for econometric estimation that accounts for the participation selection bias problem. Unfortunately there is a second type of selection bias that is likely to influence the determinants of nurses' earnings – occupation selection bias. We consider this in the next section.

4.8. Occupation selection bias

We wish to know the returns to being employed as a nurse. Specifically we wish to measure the following:

$$E(\ln W_i \mid \text{employed as a nurse}) - E(\ln W_i \mid \text{employed in an occupation other than nursing}) \qquad [4.33]$$

This may be estimated empirically by comparing the actual earnings of nurses and the actual earnings of individuals employed in alternative occupations (for example as conducted in Chapters 2 and 3). Unfortunately however as noted in Chapter 3, this may overstate or understate the true returns to being employed as a nurse depending on whether an individual employed as a nurse would earn higher, lower or the same wages if employed in another occupation than someone already employed in that occupation and on whether an individual employed in another occupation would earn higher, lower or the same wages if employed as a nurse than someone already employed as a nurse. For example, a nurse may have very limited earnings potential if otherwise employed as an accountant, but by the same token an accountant may have a much lower earnings potential as a nurse than those who actually end up choosing that kind of work.

The direction of the occupation selection bias and its influence on the estimated returns to being employed as a nurse depends on the distribution of innate ability and motivation in the population and the role of these in determining earnings in a specific occupation. Based on Willis (1986) we describe five possible patterns of selection bias that might occur in this context, derived from the notion of comparative advantage. For the purposes of the exposition we will assume for now that individuals employed in occupation n (nursing) earn a greater income than individuals employed in occupation o (some other alternative occupation).

Case 1. No occupation selection bias

An individual x employed in occupation n would earn the same wage as an individual y employed in occupation o if individual x were also employed in occupation o. Additionally, individual y employed in occupation o would earn the same wage as individual x employed in occupation n if individual y were also employed in occupation n. In this case there is no selection bias.

Case 2. Positive hierarchical sorting

Individual x employed in occupation n would earn a higher wage than individual y employed in occupation o if individual x were also employed in occupation o. Additionally, individual y employed in occupation o would earn a lower wage than individual x employed in occupation n if individual y were also employed in occupation n. In simple terms, those with universally the most ability will choose the most well paid jobs. In this case the returns to being a nurse would be overestimated.

Case 3. Negative hierarchical sorting

Individual x employed in occupation n would earn a lower wage than individual y employed in occupation o if individual x were also employed in occupation o. Additionally, individual y employed in occupation o would earn a higher wage than individual x employed in occupation n if individual y were also employed in occupation n. In simple terms, those with universally the most ability will choose the least well paid jobs. In this case the returns to being a nurse would be underestimated.

Case 4. Non-hierarchical sorting type 1

Individual x employed in occupation n would earn a lower wage than individual y employed in occupation o if individual x were employed in occupation o. Additionally, individual y employed in occupation o would earn a lower wage than individual x employed in occupation n if individual y were employed in occupation n. In simple terms those who are best at n would choose n and those who are best at o would choose o. In this case the effect on the returns to being a nurse are ambiguous.

Case 5. Non-hierarchical sorting type 2

Individual x employed in occupation n would earn a higher wage than individual y employed in occupation o if individual x were employed in occupation o. Additionally, individual y employed in occupation o would earn a higher wage than individual x employed in occupation n if individual y were employed in occupation n. In simple terms those who are worst at n would choose n and those who are worst at o would choose o. In this case the effect on the returns to being a nurse are ambiguous.

We can examine these issues more formally in a form suitable for econometric estimation. Assume first that individual i participates in the labour market (for now we dissolve the problem of participation selection bias). Let τ_{ni} be individual i's utility if they choose to be a employed as a nurse and τ_{oi} be their utility if they choose to be employed in some occupation other than nursing, where:

$$\tau_{ni} = a(W_{ni}, H_{ni}, C_i, T_{ni}, r_i) \qquad [4.34]$$

$$\tau_{oi} = b(W_{oi}, H_{oi}, C_i, T_{oi}, r_i) \qquad [4.35]$$

where W represents hourly wages, H represents hours worked, C represents personal characteristics, T represents tastes for each occupation and r is the rate of discount.

Suppose also that:

$$W_{ni} = c(X_i, A_{ni}) \qquad [4.36]$$

$$W_{oi} = d(X_i, A_{oi}) \quad [4.37]$$

where X is a matrix of measurable individual productive characteristics (X will not necessarily equal C) and A represents unmeasurable factors influencing earnings potential for each occupation. A_n and A_o may be thought of as measures of exogenous innate ability relevant to being employed as a nurse and in some other occupation, respectively. It follows that:

$$\tau_{ni} = e(X_i, A_{ni}, H_{ni}, C_i, T_{ni}, r_i) \quad [4.38]$$

$$\tau_{oi} = f(X_i, A_{oi}, H_{oi}, C_i, T_{oi}, r_i) \quad [4.39]$$

The standard Mincerian earnings functions that we wish to measure separately for nurses and other workers (of the form described by equation [4.1]) are:

$$\ln W_{ni} = \beta_n X_{ni} + U_{ni} \quad [4.40]$$

for nurses, and:

$$\ln W_{oi} = \beta_o X_{oi} + U_{oi} \quad [4.41]$$

for other workers. X is defined as before, β_n and β_o are sets of vectors of parameters pertaining to nurses and all other workers, respectively, and U_n and U_o are error terms. The important point is that U_{ni} and U_{oi} denote permanent individual-specific unobserved components that reflect unmeasured factors, including innate ability, A_i, that influence earnings potential in each specific occupation. The key question is: does $\beta_n X_{ni}$ - $\beta_o X_{oi}$ estimated by OLS, measure accurately the value to being employed as a nurse? The answer, as we have seen above, is 'no' if the typical individual who chooses to be employed as a nurse would have earned higher or lower wages than someone employed in alternative occupation if the individual had not been employed as a nurse.

An individual will decide to be employed as a nurse if the present value of their lifetime utility as a nurse exceeds the present value of their lifetime utility if they are employed in another occupation that is, if $\tau_{ni} > \tau_{oi}$. Note that this decision will be affected by, among other things, the individual's productive characteristics (the X's) but also by their innate ability in different occupations (the A's).

We can delineate this model in the same framework as that used in the participation selection bias model above where an individual's propensity to be employed as a nurse or in some other occupation can therefore be measured by a continuous variable Nu*, defined as:

$$Nu_i^* = \tau_{ni} - \tau_{oi} = \gamma z_i + v_i \qquad [4.42]$$

where Nu* is an unobserved latent variable and we observe in practice only whether an individual is employed as a nurse (Nu = 1), or is employed in some other occupation (Nu = 0). When the value of Nu* is positive, the individual decides to be employed as a nurse, and vice versa, that is:

$$Nu_i = 1 \text{ if } Nu_i^* > 0 \qquad [4.43]$$

$$Nu_i = 0 \text{ if } Nu_i^* \leq 0 \qquad [4.44]$$

z is a vector of regressors, γ is a vector of parameters and v is an error term with zero mean and constant variance, σ_v^2.

Following a similar methodology to the previous model we obtain the following earnings function for nurses:

$$E(\ln W_{ni} \mid Nu = 1) = \beta_n X_{ni} + E(U_{ni} \mid Nu = 1) \qquad [4.45a]$$

$$= \beta_n X_{ni} + \rho\sigma_{U_n}\lambda(nu)_{ni} \qquad [4.45b]$$

$$= \beta_n X_{ni} + \beta_{\lambda(nu)n}\lambda(nu)_{ni} \qquad [4.45c]$$

where

$$\lambda(nu)_{ni} = \frac{\phi(\gamma z_i / \sigma_v)}{1 - \Phi(\gamma z_i / \sigma_v)} \qquad [4.46]$$

and the nu of $\lambda(nu)$ denotes an adjustment for occupation selection bias (the decision to become a nurse or not). This may be estimated using the two-step strategy proposed by Heckman (1979) discussed above in the context of participation selection bias. The first step is the probit model of occupation selection (called the occupation selection equation) and the second step is estimation of the wage equation. The result from estimating equation [4.45c] will be a different value for $\beta_n X_{ni}$ than that obtained via OLS using equation [4.1] to allow for the self-selected nature of the occupational choice decision (the effect of choosing to be employed as a nurse).

For individuals who choose not to be employed as nurses but who choose to be employed in some alternative occupation a similar adjustment is warranted. For non-nurses the following earnings function is obtained:

$$E(\ln W_{oi} \mid Nu = 0) = \beta_o X_{oi} + E(U_{ni} \mid Nu = 0) \quad [4.47a]$$

$$= \beta_o X_{oi} + \rho \sigma_{Un} \lambda(nu)_{oi} \quad [4.47b]$$

$$= \beta_o X_{oi} + \beta_{\lambda(nu)o} \lambda(nu)_{oi} \quad [4.47c]$$

where

$$\lambda(nu)_{oi} = \frac{-\phi(-\gamma z_i / \sigma_v)}{\Phi(-\gamma z_i / \sigma_v)} \quad [4.48]$$

See Willis (1986) or Greene (2000) for a discussion and derivation of equations [4.42]-[4.48].

As in the participation selection bias model the solution to the occupation selection bias problem is to include an additional regressor [λ(nu)] that corrects the market wage in a specific occupation to account for an individual's propensity to be employed in that occupation. As discussed above this formulation is suggested by the theory of comparative advantage in terms of the innate ability relevant to being employed as a nurse or in some other occupation.

As explained in greater detail below we estimate a statistical model that adjusts for this form of selection bias. In Chapter 5 the occupation selection equation in this model is estimated using the sub-sample of individuals in the data who choose to participate. Separate wage equations are estimated using the sub-sample of this group of individuals who are employed as nurses and those who are employed in occupations other than nursing.

4.9. Selection bias correction term (λ)

4.9.1. Interpreting λ in the participation selection bias model

The following is based on Killingsworth (1983). Wages are usually not observed at all for individuals who do not work. This is problematic if we wish to delineate the relationship between a set of human capital characteristics X and wages lnW in the following statistical model:

$$\ln W_i = \beta X_i + U_i \qquad [4.49]$$

A simple method of dealing with this problem is to derive an 'imputed wage' for non-workers by assuming that the wage equation for workers and non-workers is the same and using parameter β to compute an imputed wage for non-workers. However, as the above discussion suggests there is a potential participation selection bias problem here since the error term U_i in equation [4.31] may not be a mean-zero random variable in the population as a whole. For example, it is possible that individuals who work are individuals with either above average or below average values of U. To correct for this problem we instead estimate the wage equation using the Heckman two-step procedure described above. The resulting wage equation is that given in equation [4.50]:

$$\ln W_i = \beta X_i + \beta_{\lambda(p)}\lambda(p)_i + \varepsilon_i \qquad [4.50]$$

Compared with the earlier OLS model (equation [4.49]) the βX_i in equation [4.50] has a slightly different interpretation. βX_i in equation [4.50] is an unbiased estimate of the (unobserved) wage that an individual in the population as a whole (including workers and non-workers) with productive characteristics X can earn, on average. $\beta X_i + \beta_{\lambda(p)}\lambda(p)_i$ on the other hand is an unbiased estimate of the (observed) wage that a worker with productive characteristics X can earn, on average. Estimation of equation [4.50] therefore allows wages for non-workers to be imputed, while correcting at the same time for potential participation selection bias.

In this model, the selection bias correction term [$\lambda(p)$] may be unambiguously interpreted as measuring the effect of choosing to participate. Since $\lambda(p)$ is positive a positive value of $\beta_{\lambda(p)}$ (its co-efficient) is interpreted to mean that individuals who participate have a higher expected value of lnW than those who choose not to participate, and vice versa. In summary, $\lambda(p)$ in the participation selection bias model may be interpreted as follows:

1. If the co-efficient on $\lambda(p)$ is positive, individuals who participate will earn a higher expected wage than (the same) individuals who do not participate would earn if they chose to participate.
2. If the co-efficient on $\lambda(p)$ is negative, individuals who participate will earn a lower expected wage than (the same) individuals who do not participate would earn if they chose to participate.

4.9.2. Interpreting λ in the occupation selection bias model

The following is based on Dolton and Makepeace (1987). As in the above case economists frequently interpret the selection bias correction term (λ) as the effect of making a particular choice. So, for example, the term $\beta_{\lambda(t)n}\lambda(nu)_{ni}$ in equation [4.45c] would be interpreted as measuring the effect of choosing to become a nurse. Since $\lambda(nu)_n$ is positive a negative value for $\beta_{\lambda(nu)n}$ (its co-efficient) might be interpreted to mean that individuals who choose to become nurses have a lower expected value of $\ln W_n$ than those who choose to be employed in alternative occupations if they chose to become nurses instead. In this way positive values for $\beta_{\lambda(nu)n}$ and $\beta_{\lambda(nu)o}$ in equations [4.45c] and [4.47c], respectively, are often taken as evidence in favour of the theory of comparative advantage. The basis for this conclusion is to take the expectation of earnings for individuals who choose to become nurses, say, as (the following is derived from equation [4.45c]):

$$E(\ln W_{ni} \mid \ln W_{ni} \text{ is observed}) = \beta X_i + \beta_{\lambda(nu)n}\lambda(nu)_{ni} \quad [4.51a]$$

$$= E(\ln W_{ni}) + \beta_{\lambda(nu)n}\lambda(nu)_{ni} \quad [4.51b]$$

From equation [4.51b] the selection bias effect is the difference between the conditional expected value of $\ln W_{ni}$ given employment as a nurse and the unconditional expected value of $\ln W_{ni}$. If, say, $\beta_{\lambda(t)n}$ is negative, the expected value of $\ln W_{ni}$ is lower if the individual is treated as self-selected nurse than if the individual is treated as a member of the population as a whole. In such situations it is natural to argue that individuals who choose to become nurses are of lower earnings potential than individuals in other occupations if they choose to become nurses. This interpretation is similar to that of $\lambda(p)$ in the participation selection bias model.

Unfortunately, as pointed out by Dolton and Makepeace (1987) there are difficulties with this interpretation in the occupation selection bias model, and such an unambiguous interpretation of λ is not possible. Intuitively this is because $\lambda(nu)_n$ does not include the full correction for selection bias because it does not account for the self-selected nature of the decision to be employed in an occupation other than nursing (the alternative choice).

More formally, from the above discussion we have the following occupation selection bias model:

Occupation selection equation:

$$Nu_i^* = \gamma z_i + v_i$$

$$Nu_i = 1 \text{ if } Nu^*_i > 0 \text{ and } Nu_i = 0 \text{ if } Nu^*_i \leq 0$$

$$\text{Prob}(Nu_i = 1) = \Phi(\gamma z_i) \text{ and } \text{Prob}(Nu_i = 2) = 1\text{-}\Phi(\gamma z_i)$$

Wage equation: $\ln W_{ni} = \beta_n X_{ni} + U_{ni}$, observed only if $Nu_i = 1$

$\ln W_{oi} = \beta_o X_{oi} + U_{oi}$, observed only if $Nu_i = 0$

In other words the occupation selection equation $Nu_i^* = \gamma_i z_i + v_i$ determines which wage equation to use. Using the notation described above we have:

$$E(\ln W_{ni} \mid Nu = 1) = \beta_n X_{ni} + \beta_{\lambda(nu)n} \lambda(nu)_{ni} \qquad [4.52]$$

$$E(\ln W_{oi} \mid Nu = 0) = \beta_o X_{oi} + \beta_{\lambda(nu)o} \lambda(nu)_{oi} \qquad [4.53]$$

When the individual is able to choose between the two sectors (occupation groups) the difference in the conditional expectation of lnW is as follows:

$$E(\ln W_{ni} \mid Nu = 1) - E(\ln W_{oi} \mid Nu = 0) = \Delta \qquad [4.54]$$

where

$$\Delta = \beta_n X_{ni} - \beta_o X_{oi} + \beta_{\lambda(nu)n} \lambda(nu)_{ni} - \beta_{\lambda(nu)o} \lambda(nu)_{oi} \qquad [4.55]$$

We can simplify equations [4.52] and [4.53] be assuming that $X_{ni} = X_{oi}$ (= X_i). Equation [4.55] now becomes:

$$\Delta = (\beta_n - \beta_o)X_i + (\beta_{\lambda(nu)n} - \beta_{\lambda(nu)o})\lambda(nu)_{ni} + \beta_{\lambda(nu)o} \frac{\lambda(nu)_{ni}}{\Phi(-\gamma z_i / \sigma_v)} \qquad [4.56]$$

where the last term on the right-hand side of equation [4.56] captures the relationship between the selection bias correction terms across occupations and is a measure of $\lambda(nu)_{oi}$ couched in terms of $\lambda(nu)_{ni}$ (see Dolton and Makepeace, 1987, for a proof). In this case the selection bias effect of the decision to become a nurse on $\ln W_{ni}$ clearly does not depend only on $\beta_{\lambda(nu)n}$, as equation [4.52] suggests. The upshot is that the interpretation of the selection bias effect on $\ln W_{ni}$ cannot be unambiguously signed given the sign of $\beta_{\lambda(nu)n}$.

In summary, in the occupation selection bias model we conclude that the interpretation of the selection bias effect given only the sign of the co-efficient on the selection bias correction term $\lambda(nu)$ is ambiguous.

4.9.3. Marginal effects of regressors that appear in both the participation/occupation selection equations and the wage equations

The marginal effect on lnW of variables that appear as regressors in both the participation/occupation selection equation and the wage equation consists of two components. There is the direct effect on the mean of lnW, which is the co-efficient β in the wage equation. In addition, for independent variables that also appear in the participation/occupation selection equation an indirect effect on lnW will also be exerted through their influence on λ.

Given equation [4.50] in the participation selection bias model, the marginal effect (direct plus indirect) of changes in a regressor (denoted by subscript k) that appears in both the participation equation and the wage equation is given by:

$$\frac{\delta[E(\ln W_i)]}{\delta X_{ki}} = \beta_k + \frac{\delta[\beta_{\lambda(p)}\lambda(p)]}{\delta X_{ki}} \qquad [4.57]$$

This is the marginal effect of the regressor on workers' earnings. β_k is an unbiased estimate of the return to a productive characteristic (denoted by subscript k) that an individual in the population as a whole (including both workers and non-workers) can obtain, on average. This is the direct effect. It captures the effect of the regressor on average earnings across the whole population. The indirect effect is captured by the second term on the right-hand side of equation [4.57]. This quantifies the effect that a regressor has on the decision to participate. It is a function of the co-efficient in the participation equation [δ_k] weighted by the selection bias correction term [$\lambda(p)$] and its impact on wages [$\beta_{\lambda(p)}$]. See Dolton and Makepeace (1987) for a proof. The indirect effect applies only to individuals who work.

In the occupation selection bias model we estimate separate wage equations for nurses and all other workers. In this case the marginal effect *in the observed sample* (e.g. the sample of nurses n) of changes in a regressor (denoted by subscript m) that appears in both the occupation selection equation and the wage equation is given by:

$$\frac{\delta[E(\ln W_{ni}|Nu=1)]}{\delta X_{nmi}} = \beta_{nm} - \frac{\delta[\beta_{\lambda(nu)n}\lambda(nu)_n]}{\delta X_{nmi}} \qquad [4.58]$$

The direct effect is β_{nm} . This is an unbiased estimate of the direct effect of a particularly productive characteristic (denoted by subscript m) that an individual employed as a nurse can obtain, on average. This captures the effect of the regressor on average earnings across the selected sample of the population who are employed as nurses. β_k in equation [4.57] has a different interpretation to β_{nm} in equation [4.58]: the key difference is that β_k applies to the whole population whereas β_{nm} applies only to the self-selected sub-sample. Note that the left hand side of equation [4.57] considers the partial derivative of the unconditional expectation of lnW with respect to X_k in the whole population, whereas the left hand side of equation [4.58] considers the partial derivative of the conditional expectation of $\ln W_n$ given the decision to be employed as a nurse with respect to X_m.

Regards the indirect effect in the occupation selection bias model, the change in the probability of choosing to be employed as a nurse affects the mean of $\ln W_n$ in that the mean in the group Nu = 1 (nurses) is different to that in the group Nu = 0 (all other workers). The second term in equation [4.58] compensates for this effect leaving only the marginal effect of a change given that Nu = 1 to begin with. See Greene (2000) for a proof. For example, suppose the earnings of nurses are greater than the earnings of identical workers in other occupations. Suppose also that having a degree affects positively both the probability of being employed as a nurse and earnings in either state. The marginal effect of having a degree therefore has two parts, one due to its influence in increasing the probability of entering a higher income group and one due to its influence on income within the group. In this case, the co-efficient on having a degree in the wage equation for nurses (the direct effect) will overstate the marginal effect of having a degree for nurses. The comparable co-efficient in the wage equation for all other workers will understate the marginal effect.

As both Dolton and Makepeace (1987) and Greene (2000) point out, it is quite possible that the magnitude, sign and statistical significance of the marginal effects in both the participation selection bias model and the occupation selection bias model might all be different from those of the direct effect given by the relevant co-efficient in the wage equation.

4.10. Heteroscedasticity in the Heckman two-step procedure

Using the general selection bias model given above we obtain the following selection-bias-corrected sector (e.g. occupation) equations for sectors S=1,2:

$$Y_{1i} = \beta_1 X_{1i} + \beta_{\lambda 1}\lambda_{1i} + \varepsilon_{1i} \qquad [4.59]$$

$$Y_{2i} = \beta_2 X_{2i} + \beta_{\lambda 2}\lambda_{2i} + \varepsilon_{2i} \qquad [4.60]$$

Heckman (1979) shows that resulting estimates are inefficient because the error terms (ε_{1i}, ε_{2i}) are heteroscedastic. For example, in the case of sector 1, both Puhani (2000) and Greene (2000) have shown that:

$$var(\varepsilon_{1i}) = var(U_{1i}) - \beta_{\lambda 1}^2[(\gamma z_i)\lambda_{1i} + \lambda_{1i}^2] \qquad [4.61]$$

The upshot is that $var(\varepsilon_{1i})$ is not constant but varies over i since it varies with z_i – the 'selection characteristics' of individuals in the sample (this features in the second term on the right-hand side of equation [4.61]). To correct for this problem both the standard errors on the co-efficients in the selection-bias-corrected sector equations and $var(\varepsilon_{1i})$ may be adjusted. The technical details of the correction are given in Greene (1995) and Greene (2000). Modern econometrics packages correct for this problem automatically. Suffice it to say that while heteroscedasticity is introduced with the use of the basic Heckman two-step procedure the estimates in this analysis are appropriately corrected for this.

4.11. The full information maximum likelihood (FIML) model

The parameters of Heckman's procedure for correcting for selection bias can also be estimated by a simultaneous equation full information maximum likelihood (FIML) system as an alternative to the two-step method described above. The FIML method is now described briefly. Suppose we have the following two-equation model:

Selection mechanism: $P_i^* = \delta Z_i + V_i$ [4.62]

Regression model: $\ln W_i = \beta X_i + U_i$ [4.63]

$\ln W$ is observed only if $P_i^* > 0$. Suppose that U_i and V_i have a bivariate normal distribution with zero means, constant variances σ_U^2 and σ_V^2, respectively, and correlation ρ. The FIML method seeks to estimate equations [4.62] and [4.63] jointly by maximum likelihood rather than using maximum likelihood to first estimate the probit selection mechanism model and then estimate the regression model by OLS. The likelihood function L of the model described by equations [4.62] and [4.63] can be written as:

$$L = \prod_{P^*_i \le 0} 1 - \Phi\left(\frac{\delta Z}{\sigma_V}\right) \prod_{P^*_i > 0} \Phi\left[\left(\delta Z + \frac{\rho}{\sigma_U^2}(\ln W - \beta X)\right) \Big/ \sqrt{\sigma_V^2 - \frac{\rho^2}{\sigma_U^2}}\right] * \frac{1}{\sigma_U}\phi\left(\frac{\ln W - \beta X}{\sigma_U}\right) \quad [4.64]$$

See Amemiya (1985) for a proof. Intuitively this procedure is different to the two-step method (which is also sometimes called the limited information maximum likelihood [LIML] method) because it consists of the joint estimation of the selection mechanism (the participation selection equation or the occupation selection equation) and the regression model (the wage equation) by maximum likelihood. The main difference between the two methods is in terms of the estimation procedure, as follows:

<u>Two-step model</u>: (1) Estimate the selection mechanism (the participation selection equation or occupation selection equation) as a probit model by maximum likelihood; then,

(2) Estimate separately the regression model (including the selection bias correction term) (the wage equation) by OLS.

FIML model: Estimate the selection mechanism and the regression model simultaneously by maximum likelihood.

While the FIML method is a consistent and efficient way of controlling for the selection bias problem (Greene, 1995) we opt for the two-step method here for two reasons. First, when the model is estimated by the FIML method there is no selection bias correction term (λ) in the resulting equation, though λ may be computed indirectly by other means. Second and more importantly, the wage equation in the two-step procedure is estimated by OLS, which means that the estimated fit passes through the sample mean (that is, $\ln \overline{W} = \hat{\beta}\overline{X} + \hat{\beta}_{\lambda}\lambda$ when β and β_{λ} are estimated by OLS). This is a necessary condition for analysing by decomposition analysis the nature of the observed wage differential between nurses and other workers, which is an integral part of the empirical analysis. [12] The condition does not hold for maximum likelihood estimation.

The upshot is that it is not possible to investigate fully the wage differentials between nurses and other workers using the FIML approach, which is one of the aims of the thesis. Additionally, the FIML method has been found to be quite cumbersome (Greene, 2000), and until very recently took a lot of computing power and time (Puhani, 2000). Hence, use of this method in empirical work is relatively uncommon (Vella, 1998).

4.12. A critique of the Heckman correction for selection bias

The Heckman two-step procedure outlined above provides an elaborate but feasible way of correcting problems of selection bias, and full use of this method shall be made in the next two chapters to estimate earnings functions for nurses in the NHS. Unfortunately the procedure is not without its critics. In fact it has been criticised on three grounds: the robustness of results and exclusion restrictions; the bivariate normality assumption; and, the numerical importance of the results. The following discussion is based on the participation selection bias model (it could equally apply to the occupation selection bias model). To reiterate, suppose we have the following two-stage model:

Selection mechanism (participation equation): $P_i^* = \delta_i Z_i + V_i$ [4.65]

Regression model (wage equation): $\ln W_i = \beta X_i + \beta_{\lambda(p)}\lambda(p)_i + \varepsilon_I$ [4.66]

4.12.1. Robustness of results and exclusion restrictions

Z, and therefore the selection bias correction term λ, is often highly collinear with X. This means that estimates of β tend to be unstable, non-robust and sensitive to minor changes in model specification. For example, Puhani (2000) and Moffitt (1999) both report results of surveys of the literature that show that the standard errors on β can by very high if the degree of collinearity between Z (and therefore λ) and X is high. The lack of exclusion restrictions is one likely reason for these collinearity problems. In empirical estimations, X and Z often have variables in common. In some cases the variables in each might even be identical. In this case there are no exclusion restrictions because no variables that are in X are excluded from Z. This leads to identification problems with the regression model. The result is that collinearity problems are likely to arise. As noted by Puhani (2000): "For the empirical analysis of, say, wage equations, the standard procedure to solve collinearity problems would

[12] See Chapter 5 section 5.3.

be to find appropriate exclusion restrictions. That is to say, one has to find variables that determine the probability to work (selection equation), but not the wage rate (outcome equation) directly. Practical examples of such variables could be the income of the spouse, household wealth, non-labour household income, or children (especially when estimating female wage rates)... If the empirical researcher is not able to solve the collinearity problem, the advice to be drawn from the studies surveyed here would be to use standard OLS to estimate empirical wage equations." In the analysis in the next two chapters we do address this issue. The wage equation is identified by including variables as regressors in the selection mechanism that are good (i.e. statistically significant) predictors of the dependent variable but are not associated with wages when other covariates are controlled.

4.12.2. The normality assumption

A second problem relates to the joint normality assumption. As described above an important underlying assumption in the Heckman two-step procedure is that the error terms in the selection mechanism and the regression model (V and U in the above case) are bivariate normal distributed. One criticism of Heckman's procedure is that the distributional assumption of bivariate normality is unwarranted and may be false. It has been argued that identification of the model is made on the basis of an arbitrary distributional assumption (see Greene, 2000, and Moffitt, 1999), especially if there are no exclusion restrictions.

This issue has been addressed by the developing semi-parametric literature on selection bias models (see, for example, Vella, 1998, for a review). The basic concept is that first-stage equations for the selection mechanism be obtained from semi-parametric or non-parametric methods, thus reducing the need for parametric assumptions. These approaches have the virtue of greater generality. Unfortunately, as pointed out by Greene (2000), the cost is that they generally are quite limited in the breadth of the models they can accommodate. For example, the non-parametric approach suggested by Manski (1990) is defined only for two regressors. Additionally, to date the new methods have been very little used in empirical analyses (Moffitt, 1999), and therefore their potential in addressing the potential difficulties associated with the Heckman two-step model have yet to be assessed. The outcome is that the estimation issue remains unsettled. There are no strong a priori arguments for the adoption of any of the approaches but on practical grounds Heckman's original model built around the bivariate normal distribution retains its attractiveness.

4.12.3. Numerical importance of the adjustment

A third argument often given as a criticism of the Heckman correction for selection bias is that adjustment for selection bias does not matter in any case (Moffitt, 1999). This view is in conflict somewhat with the first two issues, which imply that if collinearity is high between the selection mechanism and the regression model, or if the bivariate normal distribution assumption if false, the estimates from the model are not capable of leading to a conclusion one way or the other on the importance of the selection bias problem. In the next two chapters we present empirical earnings functions estimates based on both the unadjusted OLS estimates and selection bias corrected estimates to allow a comparison between the two views.

4.13. Earnings functions for nurses: the literature

Following the work of Mincer (1974) and Heckman (1979) outlined above we now have a theoretical basis with which to analyse the determinants of nurses' earnings. We now review the literature to date on earnings functions for nurses. We do this to examine the methodologies employed to see in particular how the work of Mincer and Heckman has been used empirically. We also wish to understand the results of earlier work to ascertain the likely determinants of nurses' earnings. In summary, we are particularly interested in finding what variables have been included in the extended earnings functions and how the problem of selection bias has been addressed. Both of these are crucial issues to the analysis in Chapters 5 and 6. In this way the previous studies will be used to inform the methodology employed.

4.13.1. Earnings functions for US nurses

The literature on earnings functions for nurses is dominated by studies from the US and based on that particular nursing labour market. These studies can be divided into four types, in all of which earnings functions for nurses appear to a greater or lesser extent, but which focus on different specific issues. The four types of study are: analyses of nursing labour supply; analyses of monopsony power in the nursing labour market; analyses of the returns to different types of nursing education; and, analyses of factors affecting the growth in wage rates of nurses over time. The separation of studies in this way is somewhat arbitrary but it is useful and informative to disaggregate the studies because it explains the specification of the earnings function that is estimated.

4.13.1.1. Studies estimating earnings functions for nurses in analyses of nursing labour supply

Most quantitative studies examining the labour supply of nurses have tried to predict how nurses would respond to increases in wages. The primary concern has been whether or not wage increases would help eliminate staff shortages, which were considered to be particularly acute in the US in the early 1970s. Clearly, if nursing labour supply were not responsive to wage increases then raising wages would simply increase labour costs, at least in the short-run.

Because the majority of nurses are female the methodology adopted in studies of nursing labour supply has been to apply general models of female labour force participation to data on nurses in order to obtain the required wage elasticities. The number of studies undertaken in this area has been fairly substantial, though the studies themselves are generally quite dated: Benham (1971); Bishop (1973); Boganno, Hixson and Jeffers (1974); Sloan and Richupan (1975); Link and Settle (1979, 1981); and, Ahlburg and Mahoney (1996) all investigate the labour supply decisions of US nurses. The general methodology has been to estimate wages using an earnings function, usually for workers only, and then to include the imputed wage in a regression of labour market and productive characteristics on hours worked.

The general results of these studies in terms of labour supply are that nurses respond positively to increases in wages, and that they work less as their property income rises. In terms of earnings function estimates for nurses, however, these studies are not at all illuminating: first; none of the studies with the exception of Ahlburg and Mahoney (1996) correct for selection bias in their earnings function estimates. Second, invariably none of the studies actually report results of the earnings functions. Therefore, while one might expect the labour supply literature for nurses to provide a rich source of data on earnings functions clearly an alternative source of data is required to review evidence on the determinants of nurses' earnings.

Note incidentally that little quantitative work has been undertaken on the labour supply behaviour of British nurses. One study has addressed this issue (Phillips, 1995), which is considered in more detail below.

4.13.1.2. Studies estimating earnings functions for nurses in analyses of monopsony power in the nursing labour market

A number of researchers have sought to test the hypothesis that monopsony power is an important determinant of wages in nursing labour markets. The earliest of these studies (Hurd, 1973, Link and Landon, 1975, Feldman and Scheffler, 1982, Bruggink et al., 1985, and Adamache and Sloan, 1982) examine the cross-sectional relationship between the average wages of nurses in different hospitals and the level of hospital concentration geographically. A further study by Sullivan (1989) estimates the monopsony power of hospitals by estimating the percentage gap between the equilibrium wage and the wage that would have prevailed had the hospital been a price taker. Most of these studies have concluded that higher concentration of employers is associated with lower mean nursing wages and that there is a monopsony effect. Unfortunately in terms of earnings functions for nurses these studies are not very forthcoming, primarily because, where reported, estimates are given for mean earnings of nurses at the hospital level, rather than for individual workers.

More useful to the present study, Hirsch and Schumacher (1995) and Schumacher and Hirsch (1997) in related analyses using the same dataset (the monthly Current Population Survey Outgoing Rotation Group conducted by the US Bureau of the Census for the years 1985 to 1993) adopt a different approach to measuring the existence of monopsony power on nurses' wages. They compare relative nurse/non-nurse wage rates in 252 geographical labour markets to examine two issues: first, whether relative nursing wage rates for nurses are lower in relatively small labour markets with a limited number of employers (see Hirsch and Schumacher, 1995); and second, why nurses working in hospitals are paid more than those working in alternative surroundings (community nurses and those employed in physicians offices, see Schumacher and Hirsch, 1997). Contrary to earlier work, and to the predictions of the standard monopsony model Hirsch and Schumacher (1995) find no evidence that the relative wages of nursing personnel are positively related to either labour market size or hospital density (number of hospitals per square mile). Schumacher and Hirsch (1997) find that hospital nurses receive a wage premium over other nurses of around 20%. They estimate that a third to a half of this advantage is due to unmeasured worker ability, and the authors conclude that the remainder of the advantage probably reflects compensating differentials for hospital disamenities such as shift work.

While it is not the main focus of either of their papers, the authors report regression estimates for the determinants of nurses' earnings. The more comprehensive analysis is reported in Hirsch and Schumacher (1995). The earnings function is specified with ln hourly wages (usual weekly earnings divided by usual hours worked per week) as the dependent variable and the following independent variables: years of schooling completed; years of potential experience (measured by age minus years of schooling minus six) and its square; and dummy variables for whether or not the nurse works in a hospital, nursing home, physician office or school (four dummy variables), union coverage, sex, race (black, other non-white or white), part-time status (less than 35 hours worker per week), employment sector (federal, state, local or private), marital status (married with spouse present, ever married without spouse present or never married), number of own children aged 17 years or below in the household (0, 1, 2, 3, or 4 or more), census region (8 dummies for 9 regions), labour market size (8 sizes with 7 dummies), and 32 quarter (time) dummies.

As hypothesised by the authors registered nurses (RNs) working in hospitals are awarded considerably higher wages than RNs with similar characteristics outside of hospitals, reflecting unmeasured skill differences and wage premia for job disamenities (primarily shift work and weekend work). RNs employed in hospitals realise a 20.1% wage premium relative to those employed in nursing homes, and a 24.7% wage premium relative to those who work in physician offices. Union-non-union wage differentials among RNs are relatively small, with only a 2.9% wage differentials between RNs covered and not covered by collective bargaining agreements. Black RNs have wage rates 10.3% lower than white RNs with similar measured characteristics. The authors view this differential to be somewhat larger than that among female workers economy-wide. The returns to years of schooling are small compared with conventional wisdom (3.3% per additional years of schooling), though the authors argue this is because they estimate only within-occupation returns and not total returns to schooling. RNs employed by the federal government, state government, and local government realise wage differentials of 3.9%, 2.4% and –2.2%, respectively, relative to RNs employed by private sector employers. Male RNs realise wage rates only marginally different from female RNs (a 1.6% difference). Also, there are no penalties associated with part-time work, with part-time RNs displaying a small (2.3%) wage advantage. Female RNs with children suffer no wage disadvantage relative to male RNs and female RNs without children.

While the findings generally seem sensible the results from the earnings function should be viewed with a certain degree of caution. As we have shown above simple OLS is not the appropriate estimation technique for estimating earnings functions because it refers only to working self-selected nurses and therefore suffers from potential selection bias.

4.13.1.3. Studies estimating earnings functions for nurses in analyses of the returns to different types of nursing education

The next set of studies do in part attempt to remedy the problem of selection bias. These studies focus on estimating the returns to different types of nursing education. There are three paths to entering the nursing labour market in the US: via an associate degree – a two year degree programme offered at community colleges (AD); via a diploma in nursing, obtained in a three-year hospital-sponsored nursing school (DIP); or, via a bachelor of science in nursing (also called the baccalaureate degree) conferred by a college or university after a four-year programme (BSN). The studies reviewed in this section compare earnings for RNs across these three educational backgrounds and attempt to ascertain which avenue into the nursing profession leads to the greatest wage advantage. The study details are listed below in Table 4.2. They are reviewed chronologically here because the authors tend to build on work that preceded their own.

In terms of the general hypothesis tested, by examining the co-efficients on the relevant independent variables and controlling for other labour market and individual productive characteristics, Mennemeyer and Gaumer (1983), Booton and Lane (1985), Link (1988) and Lehrer et al. (1991) all find that BSN-qualified nurses earn higher wages than those qualified to diploma or associate degree standard. The magnitude of effect is somewhat different across studies, however. Conflicting results emerge concerning the diploma-associate degree comparison: Booton and Lane (1985) show that diploma nurses have a wage advantage over their associate degree counterparts; Mennemeyer and Gaumer (1983) and Lehrer et al. (1991) show the opposite effect; while Link (1988) shows no statistically significant difference in earnings between the two. We now examine each paper individually.

In the first paper of its kind Mennemeyer and Gaumer (1983) test the hypothesis that different routes into nursing yield different financial returns. They indeed find that the AD avenue to becoming an RN leads to a wage premium of 19% over entering via the diploma route, while for BSN preparation the premium is even greater at 57%.

In addition to nursing qualifications Mennemeyer and Gaumer (1983) find all the other independent variables to be statistical significant at conventional levels. Wages for RNs are positively related to ln years of experience in nursing (a 3% increase in earnings for every additional year of experience) if the nurse holds an administrative position (a 99% wage premium), if the nurse holds a supervisory position (6%) and if the nurse is black (43%). Working in a rural area, and undertaking duties that allow autonomous actions both have a negative impact on earnings (of –53% and –9%, respectively).

Booton and Lane (1985) build on the initial work of Mennemeyer and Gaumer (1983) by trying to ascertain why wage differentials across methods of entry into the nursing profession persist. They test whether the differential is caused by the position of employment – they hypothesise that RNs working in the hospital setting will receive lower returns to their qualifications than those employed in other settings due to the monopsonistic or oligopsonistic nature of the hospital nursing labour market. In addition to the nursing qualification dummy variables they also include interaction terms with whether or not the nurse is employed in the hospital setting. They find that nurses educated to baccalaureate degree (BSN) standard receive 7% higher wages on average than those educated to AD level, controlling for other characteristics.

Author/Year	Described sample for earnings function	Independent variables in earnings function [1]	Selection bias correction?	Technique
Mennemeyer and Gaumer (1983)	7,330 female RNs who are full-time employees	AD, BSN, [2] holds a master's degree, ln years of experience in nursing, [3] duties allow autonomous actions, holds administrative job, holds supervisory position, works in rural area, black ethnic group	No	OLS
Booton and Lane (1985)	6,442 RNs	DIP, BSN, [4] employed in hospital, interaction between AD and hospital, interaction between DIP and hospital, interaction between BSN and hospital, general duty nurse, interaction between BSN and general duty nurse, local excess demand in hospital employment, [3] years of experience as an RN, [3] experience squared, [3] age, [3] age over 40, household income, [3] children aged 6 or under in household	No	OLS
Link (1988)	20,142 RNs aged 20-64 years	DIP, BSN, [4] male, black ethnic group, years of experience as a nurse, [3] experience squared, [3] number physicians/100 population, [3] % of state workforce unionised, [3] area designated medically underserved, metropolitan area, crime/100 population, [3] 9 regional dummies, state manufacturing wage, [3] annual hours worked, [3]	Yes [5]	FIML
Lehrer et al. (1991)	30,432 RNs	Interaction between DIP and 0-4years experience, DIP and 5-9 years, DIP and 10-19 years, DIP and 20+ years, interaction between AD and 5-9 years, AD and 10-19 years, AD and 20+ years, interaction between BSN and 0-4years experience, BSN and 5-9 years, BSN and 10-19 years, BSN and 20+ years, [6] black ethnic group, Asian or Pacific Islander ethnic group. Chicago metropolitan area, non-metropolitan area	Yes [7]	FIML
Botelho et al. (1998) [8]	22,147 female RNs	Years of experience, [3] experience squared, [3] employed full-time, white ethnic group, metropolitan area	Yes [9]	Heckman two-step

[1] The dependent variable is ln hourly wage in each instance

[2] DIP is the comparison group

[3] Indicates a continuous independent variable. All other independent variables are dummy variables.

[4] AD is the comparison group

[5] Corrects for participation selection bias. The participation equation is estimated from the sample of all RNs, including those who are out of the workforce. The dependent variable is whether or not the nurse had labour market earnings. The independent variables are: DIP, BSN, [4] male, black ethnic group, age, [3] age squared, [3] area designated medically underserved, metropolitan area, 9 regional dummies, interaction between regional dummies and metropolitan area, married, number of children, [3] children at school, family income, [3] disabled, spouse disabled, females as proportion of the area labour workforce [3]

[6] Interaction term between AD and 0-4 years of experience is the comparison group

[7] Corrects for participation selection bias. The participation equation is estimated from the sample of all RNs, including those who are out of the workforce. The dependent variable is whether or not the nurse participates. The independent variables are: DIP, BSN, [4] years of experience, married, family income, children present, age 60-65 years, black ethnic group, Asian or Pacific Islander ethnic group. Chicago metropolitan area, non-metropolitan area

[8] Botelho et al. (1998) estimate separate earnings functions for RNs with AD, DIP and BSN qualifications

[9] Corrects for participation selection bias. The participation equation is estimated from the sample of all RNs, including those who are out of the workforce. The dependent variable is whether or not the nurse participates. The independent variables are: age, [3] white ethnic group, married, widowed, divorced or separated, no children

Table 4.2. Characteristics of studies estimating earnings functions for nurses in analyses of the returns to different types of nursing education

For diploma-level nurses the premium is 2%. The impact of working in a hospital on wages in not statistically significant, but both working in a hospital and having a baccalaureate degree and working in a hospital and having a diploma both reduce mean earnings (by –4% and –5%, respectively). At the same time there is no statistically significant change in wages for nurses who worked in the hospital setting with associate degrees. The authors conclude from this that non-hospital employers are willing to pay a larger wage differential to RNs holding diplomas and baccalaureate degrees. They suggest that this effect is due to the oligopsonistic nature of the hospital labour market rather than due to a lack of recognition by employers of differences between different nursing qualifications. Results for the other independent variables controlled for in this analysis are not reported.

A number of shortcomings are clear from these early studies. The earnings function is likely to be mis-specified for a number of reasons. First, human capital attainment outside of nursing is not included. Years of schooling, for example, an important regressor in the basic Mincerian earnings function, is not included. This is likely to be problematic if pre-nursing education has an impact on career progression either directly or indirectly. Second, Mennemeyer and Gaumer (1983) fail to include a quadratic specification for the experience variable. The assumption is that earnings rise linearly with experience and for the reasons discussed above, this might well not be the case. Third, Booton and Lane (1985) do not include dummy variables for race and sex which are likely to have a significant impact on earnings (Mennemeyer and Gaumer, 1983, estimate that black nurses earn 43% higher wages on average than other nurses). Fourth, neither of these early analyses adjusts for selection bias. Link (1988) also investigates the financial rewards to investments in nursing education. The main contributions to the literature of this analysis are first that comparisons of the returns to education are made over time (for 1970, 1977, 1980 and 1984), and second that the problem of participation selection bias is addressed for the first time in this setting. The analysis is comprehensive and is worth considering in greater detail. Link (1988) adjusts for participation selection bias using the FIML method. The specification of the participation equation is given but the results are not reported. The results of the earnings function are presented in Table 4.3.

Variable	1970	1977	1980	1984
Constant	1.3114*	1.6849*	2.1962*	1.8101*
DIP	-0.0128	-0.0206	-0.0194	0.0025
BSN	0.0651	0.0702*	0.0706*	0.0665*
Masters	0.1101*	0.2498*	0.2569*	0.1933*
Male	-0.0890	0.1741*	0.1491*	0.0957*
Black	-0.1338	0.0163	0.0717*	0.0700*
Experience	0.0133*	0.0136*	0.0155*	0.0153*
Experience squared	-0.0003*	-0.0003*	-0.0002*	-0.0003*
Physicians/100 population	0.9615*	0.1224*	0.2102*	0.7834*
% of state workforce unionised	0.0059	0.2492*	0.1224	0.3684*
Medically underserved	-	-0.0198	0.0170	-
Metropolitan area	0.0956*	0.0811*	0.0758*	0.1182*
Crime/100 population	0.0007	0.0026	0.0001	-
Northeast region	0.0909*	0.0049	-0.0708*	-0.0858*
North central region	0.0018	-0.0384*	-0.0623*	-0.0446*
West region	0.0522	-0.0688*	0.0290	0.0277
Connecticut	-	0.0894*	0.0183	-0.0429
New York	-	0.0408	0.1148*	-0.1632*
New Jersey	-	0.0356	-0.0420	-0.0735*
Washington	-	0.0921*	-0.0549	-
Wisconsin	-	-0.0252	-0.0879	-
Maryland	-	0.0448	0.0068	-0.0782*
Massachusetts	-	0.1535*	-0.0082	-0.1095*
State manufacturing wage	0.0179	0.0142*	0.0225*	0.0074
Annual hours	0.0003*	0.0001*	-0.0002*	0.00004
ρ	0.2669*	-0.1144	-0.4280*	-0.2862*
σ	0.5390	0.2690	0.3420	0.2870
N	3,203	3,604	3,692	4,346
Mean lnW	2.493	2.318	2.305	2.371
% RNs participating	63.9	70.4	77.1	83.3

* Statistically significant at the 95% level
Source: Link (1988)

Table 4.3. Determinants of ln hourly wages for RNs, 1970-1984

As evidenced by the insignificant co-efficient on the diploma variable in the regressions for each of the four years the author concluded that the market viewed diploma and AD nurses (the comparison group) as homogenous. BSN nurses received a wage premium over the AD group. The returns to being educated to the masters' level were more substantial. Link (1988) also finds that both male nurses and black nurses made substantial losses over females and non-blacks, respectively in 1970, but were making considerable gains in 1977 which continued to 1984. The co-efficients on years of experience and experience squared were of the expected signs. Physicians per hundred population, a proxy for the derived demand for RNs, had a positive co-efficient. RNs residing in metropolitan areas earned higher wages. Variables were created to indicate residence in US states alleged to have stringent cost containment programmes (the regional dummies). In 1997 there was little evidence that these programmes were holding down RN wages. By 1984, when all the co-efficients but one (Connecticut) are negative and significant the impacts of the cost containment programmes appear stronger. The 1977 and 1980 data allowed the opportunity to estimate the effects on RN wages associated with working in a medically underserved area. Wages in these places are not statistically significantly different from those paid in adequately served areas. The author concludes from this that lack of salary differentials in medically underserved locations helps to explain why shortages in these areas are likely to persist.

In terms of the participation selection bias correction terms, what is presented is ρ and σ separately and not the co-efficient on λ, namely β_λ. However, since $\beta_\lambda = \rho\sigma$ (see equations [4.28d]-[4.28e]), and σ is a constant, the interpretation on the statistical significance and sign of ρ is the same as β_λ. ρ, the correlation between the error terms in the participation equation and the wage equation, was statistically significant and positive in 1970 and negative in 1977, 1980 and 1984. Therefore in 1970, RNs who worked had unmeasured characteristics causing them to earn higher wages than they would have earned by those out of the labour force if those non-workers had worked. The pattern reversed by 1977.

Lehrer et al. (1991), in a similar study to those of Mennemeyer and Gaumer (1983) and Booton and Lane (1985), also estimate an earnings function for RNs to determine the wage premium to different educational backgrounds. The main contribution of this paper is that the specification of the earnings function allows for possible differences by nursing qualification in the influence of experience on wages. Non-linearities in the experience-earnings profile, rather than be specified by a quadratic term on years of experience, are instead accounted for by including dummy variables for years of experience segments: 0-4 years; 5-9 years; 10-19 years; and, 20 years and over. The analysis adjusts for participation selection bias using the FIML method. The specification of the participation equation is given but the results are not reported.

Wage differences by nurse education among recent entrants to the profession (with 0-4 years of experience) are small. The difference between BSN and diploma nurses is not statistically significant, and while significant, the difference between BSN and AD nurses is small. Within each education category, increases in wages with experience are statistically significant. The wage premium earned by baccalaureate degree nurses increases over time. As they advance from the lowest to the highest experience group their wage premia relative to AD nurses with 0-4 years of experience are 2.5%, 15,4%, 19.1% and 26.8%. These premia are smaller for RNs with other education credentials (associate degree and diploma nurses in the highest experience groups earn 16.4% and 16.8% higher wages than AD nurses with 0-4 years of experience). Black nurses and those of Asian or Pacific Islander origin earn more than white nurses (by, on average, 6% and 9%, respectively). The authors also found that nurses working in the Chicago metropolitan area earned 7% higher wages than those working outside of this area on average, while those working in rural areas earned 10% lower wages.

As in the study by Link (1988), ρ and σ are presented separately as the participation selection bias control variables. ρ is statistically significant and negative. RNs who worked had unmeasured characteristics causing them to earn lower wages than they would have earned by those out of the labour force if those non-workers had worked. Lehrer et al. (1991) also add that because ρ is small in magnitude (-0.1565), and because OLS-estimated earnings function with the same variables yield virtually the same results (not shown), selection bias "does not appear to be an important factor".

The studies reviewed so far in this section generally assume that the effect of the controlling variables on wages are the same for the three different avenues into nursing. Since the method of entry into the profession is included in the models only as a dummy variable this affects only the intercept term rather than the slope co-efficients. A different approach is adopted by the final paper reviewed in this section (Botelho et al., 1998) who estimate separate earnings functions for nurses with the three different educational qualifications. They also correct for participation selection bias in their estimates using the Heckman two-step procedure. Note that the authors also correct for a second form of potential selection bias, which in this case arises from the choice of educational credential obtained (BSN, AD or DIP). This kind of double-selectivity model is considered in more detail in Chapter 6 and this component of the analysis by Botelho et al. (1998) will be reviewed in more detail then. The results of the participation-selection-bias-corrected model are presented in Table 4.4. From the participation equation results marital status is a statistically significant determinant of the probability of participation in the labour force: the co-efficient is negative and so married females are less likely to participate than their single counterparts. Being widowed, divorced or separated exerts upward pressure on the likelihood of participation. The co-efficient on 'No children' is not statistically significantly different from zero at conventional levels. The probability of participating decreases with age and is lower for white nurses.

Variable	Co-efficient		
Participation equation			
Constant	3.5210*		
Married	-0.1403*		
Widowed, divorced or separated	0.1712*		
No children	0.0036		
White ethnic group	-0.1976*		
Age	-0.0498*		
Log-likelihood	-8311.46		
χ^2 statistic	3268.06		
Variable		Co-efficients	
Wage equations	AD	DIP	BSN
Intercept	2.6935*	2.7940*	2.7060*
Years of experience	0.0224*	0.0025	0.0139*
Years of experience squared	-0.0006*	-0.0002*	-0.0004*
Employed full-time	-0.0286*	0.0369*	0.0141
White ethnic group	-0.0889*	-0.0925*	-0.0931*
Metropolitan area	0.1599*	0.1111*	0.1619*
λ	-0.1059*	0.2550*	0.2649*
R^2	0.11	0.36	0.29

* Statistically significant at the 95% level
Source: Botelho et al. (1998)

Table 4.4. Estimates of participation selection bias-corrected wage equations for AD, DIP and BSN nurses

Turning now to the results of the participation-bias-corrected wage equations, there are a number of similarities between the determinants of ln hourly wages for nurses with different education credentials. In all three instances the experience-earnings profile is as predicted by the basic Mincerian model, namely n-shaped. Also the co-efficients on the white ethnic group and residing in a metropolitan area dummy variables are of the same sign and order of magnitude. The co-efficient on white ethnic group is negative in all three equations – that is, regardless of avenue into the nursing profession, white nurses earn less than their non-white counterparts. This is consistent with the results of Mennemeyer and Gaumer (1983), Link (1988) and Lehrer et al. (1991) who all find that nurses from non-white ethnic groups earn higher wages than white nurses after controlling for differences in educational attainment and years of nursing experience.

The wage premium on living in a metropolitan area is 11-16%. The main differences are in terms of the impact of working full-time and the impact of participation selection bias. Diploma nurses who work full-time have a wage advantage over their part-time colleagues; for associate degree nurses there is an opposite effect; for BSN nurses the relationship is not statistically significant.

In all instances λ, the participation selection bias correction term is statistically significant. For AD nurses the co-efficient is negative, while for DIP and BSN nurses it is positive. The interpretation of this result is that AD nurses who worked had unmeasured characteristics causing them to earn lower wages than would have been earned by those out of the labour force if those non-workers had worked. The opposite is true for diploma and baccalaureate degree nurses.

An important criticism of this study is that while a correction is made for participation selection bias, the participation equation is mis-specified: an important explanatory variable in the reservation wage based on the theory of labour supply is the amount of property income received, and this has been included in the analysis. The applicability of the results to the UK also is limited because of cross-country differences in methods of entry into the nursing profession. Nonetheless, this study remains a useful application of earnings functions to the nursing labour market.

4.13.1.4. Studies estimating earnings function to find factors affecting the growth in wage rates of nurses over time

We consider finally in this section a single US-based study that focuses on finding the determinants of growth in the wage rates of nurses over time. In what is the most comprehensive analysis (in terms of independent variables included) of nurses' earnings in the US to date Ault and Rutman (1998) using data from two surveys of 2,584 RNs in 1981 and 1989 examine the factors affecting the levels and growth rates in the wage rates paid to female RNs. The first stage of the analysis is to find the factors affecting observed wage levels. The second stage focuses on the factors affecting the observed growth in wage rates since the individual was first employed as a nurse. The following equations were estimated:

Participation = f({HOUSE}, {EDUC}, {WORKHIS}, {SPOUSE}) [4.67]

lnW = g({HOUSE}, {EDUC}, {WORKHIS}, {POSITION}, {EMPLOYER}, {SPEC}, {SPOUSE}, {REGION}) [4.68]

Growth = h({HOUSE}, {EDUC}, {WORKHIS}, {POSITION}, {EMPLOYER}, {SPEC}, {SPOUSE}, {REGION}) [4.69]

In the first stage of the analysis an earnings function is estimated for RNs. To control for participation selection bias the FIML approach was used. The specification of the participation equation is given in equation [4.67]. The dependent variable was whether or not wages were observed. This was estimated from the whole sample of working and non-working RNs. The wage equation, specified in equation [4.68] was estimated for working

RNs only with the correction for participation selection bias. The second stage of the analysis was to determine the factors affecting the growth in wage rates since the individual was first employed as a nurse. The growth rates were analysed using OLS estimation of equation [4.69].

The estimated equations are composed of sets of continuous and binary variables that are designed to measure the influence of household characteristics {HOUSE}, education {EDUC}, work history {WORKHIS}, the title of the position held within the organisation where the individual works {POSITION}, the type of employer {EMPLOYER}, the individual's nursing specialisation {SPEC}, the spouse's occupation and employment status {SPOUSE}, and region of employment {REGION}. The results for the participation equation are presented in Table 4.5. The relationship between non-nursing income (as a proxy for property income) and labour force participation is not statistically significant. This is unexpected given the model of individual labour supply described above. Another surprising result is the negative co-efficient on having a masters' degree. The implication is that the more educated are less likely to participate in the labour market. These findings are not discussed or justified by the authors. The other results are generally of the expected sign. Regards the effect of children at home, for instance, in 1981 those with children in high school (aged 14-18 years) were more likely to work than those with younger or no children. In 1989 those with a child less than 6 years of age or with two children between the ages of 6 and 13 were less likely to work than those with older children or no children at home. By 1989 however the ages and number of children appear generally to have become relatively unimportant.

Variable	1981 [1]	1989 [1]
Constant	2.64***	1.39***
Married	0.25*	0.18
Age [2]	-0.040	0.0018
Age less than 40 years and no children	0.18	-0.24
Age more than 40 years and no children	0.44	-0.084
Has a child less than 1 year old	-0.23	-0.12
Has 1 child less than 6 years old	-0.092	-0.35**
Has 2 children less than 6 years old	-0.34	-0.12
Has 1 child between 6 and 13 years old	-0.055	-0.017
Has 2 children between 6 and 13 years old	0.12	-0.27*
Has 1 child between 14 and 18 years old	0.42*	-0.083
Has 2 children between 14 and 18 years old	0.83***	0.10
Household annual non-nursing income US$5,000-10,000	0.45	0.74**
Household annual non-nursing income US$10,000-25,000	0.11	0.17
Household annual non-nursing income US$25,000-40,000	-0.19	0.074
Household annual non-nursing income US$40,000-50,000	-0.081	0.30
Household annual non-nursing income US$50,000+	-0.22	-0.063
Masters degree	-0.19	-0.39***
% of years since first licensed worked part-time [2]	0.56**	-0.68***
% of years since first licensed unemployed [2]	-1.71***	-2.93***
Number of switches in employment state [2]	-0.061	0.0055
Spouse employed part-time	-0.36	-0.26
Spouse unemployed	0.23	-0.39*
Spouse employed as a medical doctor	-0.88***	-0.65***
Spouse holds professional certification	0.0036	-0.0017
Log-likelihood	-254.70	-393.53
N	1,276	1,308

[1] The dependent variable is whether or not wages were observed
[2] Indicates a continuous independent variable. All other independent variables are dummy variables
* Statistically significant at the 90% level
** Statistically significant at the 95% level
*** Statistically significant at the 99% level
Source: Ault and Rutman (1998)

Table 4.5. Results of participation equation: 1981 and 1989

The results for the wage equations are presented in Table 4.6. We discuss these in conjunction with the growth rate equations, which are presented in Table 4.7. In contrast to the results of earlier research discussed above Ault and Rutman (1998) find that wage rates do not vary significantly with educational background (Link, 1988, for example reports a positive return to the baccalaureate degree in nursing): the wage rates paid to nurses who earned associate degrees (the comparison group in the analysis) were not significantly different from those earned by individuals who had earned diplomas and baccalaureate degrees.

Variable	1981 [1]	1989 [1]
Constant	2.14***	2.60***
Married	0.023	-0.32
Number of children 0-5 years [2]	0.023	0.0096
Number children 6-13 years [2]	0.014	0.020
Number of children 14-18 years [2]	0.017	-0.0062
Household annual non-nursing income US$5,000-10,000	0.0087	-0.010
Household annual non-nursing income US$10,000-25,000	-0.0068	0.011
Household annual non-nursing income US$25,000-40,000	0.0013	0.023
Household annual non-nursing income US$40,000-50,000	-0.013	0.043
Household annual non-nursing income US$50,000+	-0.0045	0.066
DIP	-0.0034	0.015
BSN	0.0088	0.024
Non-nursing baccalaureate degree	0.026	0.0023
Masters degree	0.22***	0.13***
Employed as staff nurse	-0.070***	-0.033
Employed as head nurse	0.030	0.059
Employed as supervisor or administrator	0.12***	0.15***
Employed as nurse educator	0.079***	0.025
Employed in a hospital or clinic	0.23***	0.12***
Employed by nursing home	0.0005	-0.012
Qualified to work in intensive care unit	0.049*	0.012
Specialises in paediatrics, obstetrics or gynaecology	0.042	-0.019
Employed as a surgical nurse	0.043	0.067*
Employed in specialty other than those listed above	0.048*	0.024
Number of years licensed to practice as an RN [2]	0.0042***	0.0032***
% of years since first licensed worked part-time [2]	-0.041	-0.069*
% of years since first licensed unemployed [2]	-0.21***	-0.21**
Switched employment states once	-0.023	0.015
Switched employment states twice	-0.066**	-0.017
Switched employment states more than twice	-0.043	-0.10***
Spouse employed part-time	0.020*	0.024
Spouse unemployed	-0.071	0.023
Spouse employed as a medical doctor	-0.054	-0.12
Spouse holds professional certification	-0.026	-0.018
Spouse employed in salaried position	-0.038	0.035
Spouse compensated for employment in wages per hour	-0.025	0.0051
Spouse self-employed	-0.023	0.031
Resides in Kansas City metropolitan area	0.013	-0.011**
Resides in St. Louis metropolitan area	0.012	-0.071
ρ [2]	-0.011***	0.33***
σ	0.21	0.22
R^2	0.32	0.22
F-ratio	12.91***	7.25***
N	1,105	1,067

[1] The dependent variable is ln hourly wage
[2] Indicates a continuous independent variable. All other independent variables are dummy variables
* Statistically significant at the 90% level
** Statistically significant at the 95% level
*** Statistically significant at the 99% level
Source: Ault and Rutman (1998)

Table 4.6. Results of the analysis of observed wage levels: 1981 and 1989

Ault and Rutman (1998) justify this finding in terms of the relatively short supply of RNs in the labour market in the 1980s: competition among employers for RNs may have eliminated the premium to BSN nurses as employers bid for the services of diploma and associate degree nurses. The growth in nurses' wage rates were, however, found to be affected by the educational background. The growth in wage rates of those employed as nurses in 1989 with baccalaureate degrees (in nursing) were larger by an average of 0.8% per year since first employed as a nurse.

The wage rates of RNs who obtained degrees in fields other than nursing saw their hourly wage rates grow annually at a rate that was 1.22% lower than the growth in wages paid to those who earned associate degrees (the comparison group). These RNs consisted primarily of individuals who earned a baccalaureate degree in another field prior to switching to nursing as a career.

Ault and Rutman (1998) find that RNs with master's degrees receive significantly higher wage rates than those without this level of education. The wage growth rates however were unaffected by having a master's level degree. The authors interpret this to mean that the lifetime earnings profile of an individual who has obtained a masters' degree is shifted up relative to someone without the degree, but that the shape of the curve is unaffected.

Consistent with the Mincerian model, there is a small but positive return to each year of experience in nursing, though the omission of a quadratic term means that this relationship is likely to be mis-specified. The growth rates in wages were, however, negatively related to years of experience. The authors offer no satisfactory explanation for this finding. The effect of children at home on the levels and growth rates of hourly wages was found to be marginal.

Work history was found to have a negative impact on wages. Frequent changes in employment state as well as periods of working unemployment or part-time employment as a nurse adversely affect earnings. As the proportion of time in which the individual was out of nurse employment increased the wage rate in both 1981 and 1989 fell. The authors interpret these findings as consistent with the view that, as compared to the human capital of those who work full-time as a nurse continuously, the human capital of those who do not work at all is perceived to decay. As the results of Tables 4.8 and Table 4.9 indicate both the wage level and wage rate growth is affected negatively by switching employment states two or more times throughout a career. There is no statistically significant effect from switching employment states only once.

Variable [1]	Co-efficient [2]
Constant	6.92***
Married	-0.55
Number of children 0-5 years [3]	-0.32
Number children 6-13 years [3]	-0.62
Number of children 14-18 years [3]	-0.38
Household annual non-nursing income US$5,000-10,000	-0.13
Household annual non-nursing income US$10,000-25,000	0.63
Household annual non-nursing income US$25,000-40,000	0.97
Household annual non-nursing income US$40,000-50,000	-0.078
Household annual non-nursing income US$50,000+	-0.24
DIP	0.075
BSN	0.82**
Non-nursing baccalaureate degree	-1.22**
Masters degree	0.57
Employed as staff nurse	-1.62***
Employed as head nurse	-0.76
Employed as supervisor or administrator	-1.17
Employed as nurse educator	-0.76
Employed in a hospital or clinic	0.99**
Employed by nursing home	0.88
Qualified to work in intensive care unit	-0.23
Specialises in paediatrics, obstetrics or gynaecology	-0.94*
Employed as a surgical nurse	-2.00
Employed in specialty other than those listed above	-0.28
Number of years licensed to practice as an RN [3]	-0.19***
% of years since first licensed worked part-time [3]	1.09
% of years since first licensed unemployed [3]	3.23*
Switched employment states once	-0.0009
Switched employment states twice	-1.42***
Switched employment states more than twice	-2.07***
Spouse employed part-time	-1.00
Spouse unemployed	0.26
Spouse employed as a medical doctor	0.098
Spouse holds professional certification	-0.38
Spouse employed in salaried position	0.50
Spouse compensated for employment in wages per hour	-0.17
Spouse self-employed	0.30
Resides in Kansas City metropolitan area	0.94**
Resides in St. Louis metropolitan area	0.063
R^2	0.23
F-ratio	5.83
N	814

[1] Presumably estimated at 1989 values, though this is not stated in the paper
[2] The dependent variable is % growth in wages per year since first licensed
[3] Indicates a continuous independent variable. All other independent variables are dummy variables
* Statistically significant at the 90% level
** Statistically significant at the 95% level
*** Statistically significant at the 99% level
Source: Ault and Rutman (1998)

Table 4.7. Results of the analysis of the growth of wage rates since first employed as a nurse

4.13.2. Earnings functions for nurses in the UK

In the single application of earnings functions to the British nursing labour market to date Phillips (1995) constructs a formal labour supply model for nurses working in the NHS in Great Britain. Using 1980 data from the Women and Employment Survey on 403 females aged 16 to 59 years either currently working as nurses (nursing auxiliaries and nursing aides were also included as well as all qualified nurses) or who had previously worked as nurses but who are now out of the labour force Phillips (1995) estimates an earnings function for nurses using the Heckman two-step model to correct for participation selection bias.

The results of the initial participation equation are not presented or discussed, though to estimate the labour supply function a second participation equation is presented incorporating the imputed wages estimated from the earnings function.

The dependent variable in the earnings function is ln hourly wage. The independent variables and their sample means are listed in Table 4.8, along with the results. It is worth considering the specification of the model in greater detail. The wage equation is based on the assumption that market wages which individuals receive for their work depend on their qualifications, the level and the extent of their work experience, and the occupation and industry in which they search for a job. Many of these variables are available directly in the data. Unfortunately the Women and Employment Survey on which the analysis is based does not distinguish between different types of nursing qualifications nor length of service in the profession. Proxy variables were therefore created for these values in the wage equation. Four variables are included to capture the effects of work experience. Because nurses gain ward knowledge by changing posts relatively frequently the effect of breadth of experience is captured by a dummy variable representing whether or not the nurse has held more than five nursing posts in her career. Years in the profession also influence pay. While a pure measure of years in nursing is not available in the data a measure of general labour market experience is. In the analysis this represents the joint effect of time in nursing and time in which general skills are accumulated. As a proxy to discount non-nursing job experience a dummy variable representing whether or not the nurse has worked in two or more occupations is included. This is expected a priori to exert a negative effect on earnings.

Dummy variables for possession of nursing qualifications and different secondary qualifications are also included. The author argues that having at least one A level and having a nursing qualification are expected to have a positive wage effect because they are basic requirements for any place on the nursing promotional ladder. In the sample, holding A levels, the highest obtainable secondary school qualification, may also be an indicator of those on the 'fast track' in nursing in the sense that entry into the nursing profession may be granted without it. Two other qualifications are expected to have a negative effect on earnings. Leaving school at age 16 or younger indicates that a significant amount of human capital investment has been foregone. Having O levels as the highest academic qualification is sufficient to gain entry into the nursing profession, but alone is unlikely to increase earnings power.

Variable	Variable mean	Co-efficient
Constant		0.52*
Has obtained at least one A level	0.10	0.24*
Has a nursing qualification	0.45	0.15*
Has obtained at least one O level	0.53	-0.06
Left school at age 16 years	0.66	-0.17*
Has worked in two or more occupations	0.44	-0.26*
Has held more than 5 nursing posts	0.41	0.14*
Months of experience in the labour market [2]	157.52	0.0011*
λ		0.09
R^2		0.31
N		144
ln hourly wage	0.66	

[1] The dependent variable is ln hourly wage
[2] Indicates a continuous independent variable. All other independent variables are dummy variables
* Statistically significant at the 90% level
Source: Phillips (1995)

Table 4.8. Sample means and wage equation results for the full nurse sample

In terms of the results of the econometric model Phillips (1995) states that all variables exhibit the predicted effects. Holding a nursing qualification and A levels increase the wage, while leaving school at age 16 or younger has a negative impact. Labour market experience has a positive impact on wages, as does having held more than five nursing posts. On the other hand, having worked in two or more occupations significantly decreases earnings. The co-efficient on the selection bias correction term is not statistically significant and this is interpreted to mean that selection bias is not significant for the group.

This is a useful analysis, not least because it is the only study to date to analyse the earnings of nursing in the NHS via extended earnings functions. However, there are a number of limitations. First, in terms of its relevance to the current situation the study is now dated. In 1980 nurses' pay was determined by the Nurses and Midwives Whitley Council, a structurally different system to the present one (see Chapter 2). The late 1980s saw major changes to the structure of the nursing profession via the clinical regrading exercise.

Additionally there were significant changes to pre-registration training for nurses with the introduction of the Project 2000 reforms. These changes, along with the major NHS Reforms of 1991, mean that Phillips' estimates are unlikely to still apply. Additionally, while the study outlines possible variables to be included in an econometric model of nurses' earnings there are other variables not included that are likely to impact on wages. Obvious omissions include ethnic group, level of health and disability, experience squared, hours worked and whether or not the nurse is working full-time or part-time, and whether or not the nurse works in a metropolitan area (e.g. London) all of which have been shown in the other studies reviewed here to influence nurses' earnings. In the present analysis we are able to examine nurses' wages using a comprehensive and up-to-date dataset allowing us to further Phillips' earlier work.

4.14. Conclusions

In Chapter 4 we provide the justification and framework for the analysis of the next two chapters in which we examine the factors affecting nurses' earnings and the nature and magnitude of wage differentials between nurses and other workers. We have determined the economic model on which the analysis is based, the specification of the statistical models, the appropriate method of estimation, and the correct approach to interpreting the results.

First we constructed an economic model of earnings that is suitable for empirical estimation. The model is based on extended Mincerian earnings functions. At its simplest the model is of the form $\ln Y = \beta_0 + \beta_1 S + \beta_2 t + \beta_3 t^2$, though this is extended with the inclusion of additional exogenous variables likely to affect earnings. From the review of studies later in the chapter we found that nurses' wages depend on their qualifications, the level and the extent of their work experience, their job specification, and other personal characteristics the most prominent of which are ethnic group and geographical area of residence. These are prime candidates for inclusion in the extended model.

Second we specified the statistical model. At its simplest this was of the form $\ln W_i = \beta X_i + U_i$.

Third we considered the issue of estimation. We showed that estimation of the earnings function without allowing for the self-selected nature of the decision to participate in the labour market or to work in a specific occupation might lead to selection bias. There is no conceptually correct approach to addressing this problem but on practical grounds the appropriate estimation technique to use is the Heckman two-step procedure. Utilisation of this method leads to two further specification issues. First the specification of the statistical model is revised to the form $\ln W_i = \beta X_i + \beta_\lambda \lambda + U_i$. The second issue is to ensure that the model is identified with appropriate exclusion restrictions.

The fourth aspect we considered was the issue of interpretation – the interpretation of the co-efficients estimated by the statistical model. We noted first the correct interpretation of the selection bias correction terms. We then also interpreted by their marginal effects the co-efficients of variables that feature in both the participation/occupation selection equations and the wage equation.

Also in this chapter we have undertaken a review of the literature on earnings functions for nurses. The conclusion was that most studies to date are based on the US nursing labour market and are therefore not directly relevant to the present analysis. Moreover it is notable that these models suffer frequently from mis-specification (based on the structural models of participation and earnings) and selection bias. A single earnings function analysis for nurses in the UK is based on 1980 data and has not included a number of potentially important explanatory variables. From a methodological perspective one important finding from the literature review, which partly drives the framework for analysis outlined above, is the importance of including a correction for potential selection bias. In terms of participation selection bias as evidenced from the literature review this has been shown to be an important regressor in a number of analyses of nurses earnings. In terms of occupation selection bias the literature review has revealed that to date this issue has not been adequately addressed. The upshot is that there is a clear justification for adjusting for both these potential effects in an analysis of nurses' earnings. This will be one focus of the analysis in subsequent chapters.

In summary, given the review in this chapter it is clear that there remains a great deal of work to be done in estimating the determinants of nurses' wages in the NHS and in evaluating nurses' relative earnings. On the basis of this chapter we now have both an economic model and a statistical model with which to examine these issues.

CHAPTER 5

ANALYSIS OF EXTENDED EARNINGS FUNCTIONS FOR NURSES

5.1. Introduction

In Chapter 5 we build on the review undertaken in Chapter 4 and conduct an earnings function analysis to examine empirically the factors that affect the wages of nurses working in the NHS. We also examine the nature and magnitude of wage differentials between nurses and other workers and investigate the causes of these observed differentials. Using the extended earnings functions described previously we consider pay determination for individuals who choose to be employed as nurses and who choose to be employed in other occupations using individual data. More specifically we examine whether nurses and other workers earn comparable wages when other factors are held constant. This involves the estimation of wage equations for nurses and other workers with appropriate correction for both participation selection bias and occupation selection bias using the Heckman two-step procedure (Heckman, 1979) and then the comparison of average pay that would be received by nurses and other workers if they were paid according to the same pay schedule.

Following the review of the previous chapter we present in Chapter 5 what is essentially a new application of the Mincerian model to nurses' earnings in Great Britain. It is also worth bearing in mind that in the next chapter (Chapter 6) we build on this approach and analyse nurses' earnings with an original double selectivity framework that corrects simultaneously for two forms of selection bias. Utilising both approaches allows us to see if the results are robust across the different models. We begin by describing the statistical models to be estimated.

5.2. The statistical models

We estimate five statistical models based on the methods outlined in Chapter 4.

5.2.1. Model 1

This is a simple OLS extended earnings function including a dummy variable for whether or not an individual is employed as a nurse. The following model is estimated, based on equation [4.1]:

$$\ln W_i = \beta X_i + \beta_n Nu_i + U_i \qquad [5.1]$$

lnW is the natural logarithm of hourly wages. X is a matrix of measurable individual productive characteristics. Nu is a dummy variable indicating whether or not the individual is employed as a nurse. β is a vector of parameters, with β_n measuring the returns to being employed as a nurse. U is an error term. This model makes no adjustment for selection bias and therefore the regression co-efficients are liable to be biased. We undertake this model for comparative purposes. The model is estimated using the sub-sample of individuals in the data (described below) who participate in the labour market (i.e. workers only).

5.2.2. Model 2

Model 2 is the participation selection bias model described in Chapter 4. We construct an earnings function using the participation selection bias model including a dummy variable in the wage equation for whether or not an individual is employed as a nurse. The model is estimated using the Heckman two-step procedure, and has the following form:

Participation equation: $P_i^* = \delta Z_i + V_i$ [5.2]

Wage equation: $\ln W_i = \beta X_i + \beta_n Nu_i + \beta_{\lambda(p)}\lambda(p)_i + \varepsilon_i$ [5.3]

P* is a latent variable reflecting an individual's propensity to participate in the labour market. Z is a vector of regressors influencing labour market participation. δ is a vector of parameters. lnW, X, Nu, β and β_n have the same interpretation as in Model 1. λ(p) is included as a regressor to reflect the predicted probability of being in paid work given other known characteristics and $\beta_{\lambda(p)}$ is its coefficient. V and ε are error terms. The participation equation in this model is estimated using the whole sample of individuals in the data (workers and non-workers). The wage equation is estimated using the sub-sample of those who work.

5.2.3. Model 3

To allow for differences in slope co-efficients (the β's) between nurses and other workers in the sample Model 3 estimates separate OLS earnings functions for each sector/occupation group (i.e. nurses and all other workers). This is based on equation [4.1] in Chapter 4, which is now estimated separately for each sub-sample in two separate regressions. The two regressions are:

$\ln W_{ni} = \beta_n X_{ni} + U_{ni}$ [5.4]

$\ln W_{oi} = \beta_o X_{oi} + U_{oi}$ [5.5] where lnW, X, β and U have the same interpretation as before and the subscripts distinguish between nurses (n) and all other workers (o). The nurse dummy variable (Nu) is now omitted. This model makes no adjustment for selection bias, and again the regression co-efficients are liable to be biased. This model is also estimated for comparative purposes. As with Model 1 it is estimated using the sub-sample of individuals in the data who participate. Equation [5.4] is estimated using data for nurses only and equation [5.5] is estimated using data for all other workers only.

5.2.4. Model 4

Model 4 is a modified version of the participation selection bias model described in Model 2. It is also similar to Model 3 in that we estimate wage equations separately for nurses and other workers. However, in this case adjustments are also made to the separate wage equations for participation selection bias using the Heckman two-step procedure. For individuals employed as nurses the following model is estimated:

Participation equation: $P_{ni}^{*} = \delta Z_{ni} + V_{ni}$ [5.6]

Wage equation: $\ln W_{ni} = \beta_n X_{ni} + \beta_{\lambda(p)n}\lambda(p)_{ni} + \varepsilon_{ni}$ [5.7]

The participation equation is estimated using the sample of participating and non-participating nurses in the data (non-participating nurses are defined below). Therefore $\lambda(p)_n$ is estimated for each nurse (working and non-working) in the whole sample. The wage equation is estimated using the sub-sample of nurses who are working.

For individuals employed in occupations other than nursing the following model is estimated:

Participation equation: $P_{oi}^* = \delta Z_{oi} + V_{oi}$ [5.8]

Wage equation: $\ln W_{oi} = \beta_o X_{oi} + \beta_{\lambda(p)o}\lambda(p)_{oi} + \varepsilon_{oi}$ [5.9]

The participation equation is estimated using the whole sub-sample of non-nurses in the data (including both participating and non-participating non-nurses). $\lambda(p)_o$ is computed for each non-nurse observation. The wage equation is estimated using the sub-sample of those who are employed in occupations other than nursing.

5.2.5. Model 5

Model 5 is the occupation selection bias model described in Chapter 4. We estimate wage equations separately for nurses and other workers with adjustments to the occupation-specific wage equations for occupation selection bias. For individuals employed as nurses the following model is estimated:

Occupation selection equation: $Nu_i^* = \gamma z_i + v_i$ [5.10]

Wage equation: $\ln W_{ni} = \beta_n X_{ni} + \beta_{\lambda(nu)n}\lambda(nu)_{ni} + \varepsilon_{ni}$ [5.11]

Nu* is an unobserved latent variable reflecting whether or not the individual is employed as a nurse, z is a vector of regressors, γ is a vector of parameters and v is an error term. lnW, X, β

and ε have the same interpretation as in Model 1, though the subscript n indicates that these variables apply to nurses only. $\lambda(nu)_n$ is included as a regressor to reflect the predicted probability of being employed as a nurse given other known characteristics, and $\beta_{\lambda(nu)n}$ is its coefficient. The occupation selection equation in this model is estimated using the sub-sample of individuals in the data who participate. The wage equation is estimated using the sub-sample of this group of individuals who are employed as nurses.

For individuals employed in occupations other than nursing the following model is estimated:

Occupation selection equation: $Nu_i^* = \gamma z_i + v_i$ [5.12]

Wage equation: $\ln W_{oi} = \beta_o X_{oi} + \beta_{\lambda(nu)o}\lambda(nu)_{oi} + \varepsilon_{oi}$ [5.13]

The variables in these equations have the same interpretation as before, though the subscript o indicates that these variables apply to workers in occupations other than nursing only. The occupation selection equation in this model is estimated using the sub-sample of individuals in the data who choose to participate. [13] The wage equation is estimated using the sub-sample of this group of individuals who are employed in occupations other than nursing.

Models 3-5 are estimated to allow for differences in slope co-efficients between nurses and other workers. We estimate separate models rather than pooling observations and including dummy variables to allow for changes in slope co-efficients to allow for possible unequal

[13] Note that the occupation selection equations delineated in equations [5.10] and [5.12] will yield identical results in terms of the co-efficients γ though they will be of the opposite sign. The equations include the same variables and are estimated on the same sample in the data (all workers). They differ in that the coding for the observable binary variable Nu in equation [5.10] is 1 if the worker is employed as a nurse and 0 otherwise. In equation [5.12] these are switched and the coding is 1 if the worker is employed in an occupation other than

disturbance variances across the two groups. An additional advantage of estimating the earnings functions separately in this way is that it allows us to conduct more easily a decomposition analysis.

5.3. Decomposition analysis

Having estimated the extended earnings functions separately for nurses and other workers using Models 3-5 we may then decompose the observed pay differential between nurses and other workers into three main components: due to differences in endowments; due to differences in the returns to endowments; and (where applicable), due to differences in selection bias. The usual method, proposed by Oaxaca (1973), is to compare the average pay that would be received by workers in the two sectors if they were paid according to the same pay structure.

Based on the models outlined we either estimate the wage equations for nurses and other workers directly by OLS (in the case of Model 3) or we follow Heckman's two-step estimation procedure (for Models 4 and 5), as follows:

Two-step model: (1) Estimate the selection mechanism (the participation selection equation or occupation selection equation) as a probit model by maximum likelihood; then,

(2) Estimate separately the regression model (including the selection bias correction term) (the wage equation) by OLS.

---

nursing and 0 otherwise. Thus the co-efficients in equations [5.10] and [5.12] will be numerically identical but of the opposite sign.

The upshot is that for all three models (Models 3-5) the wage equations are ultimately estimated by OLS (with or without the adjustments for selection bias). A useful feature is that the OLS estimates pass through the sample mean. [14] We therefore estimate the following:

$$\ln \overline{W}_n = \hat{\beta}_n \overline{X}_n + \hat{\beta}_{\lambda n} \overline{\lambda}_n \qquad [5.14]$$

$$\ln \overline{W}_o = \hat{\beta}_o \overline{X}_o + \hat{\beta}_{\lambda o} \overline{\lambda}_o \qquad [5.15]$$

where the bar indicates a mean value and the hat indicates an estimated value. (Let $\ln \overline{W}$ denote the mean of the natural logarithm of wages, not the natural logarithm of mean wages.) In Model 3 the second term on the right hand side of equations [5.14] and [5.15] is omitted. Subtracting equation [5.14] from equation [5.15] to obtain the wage differential we obtain the following:

$$\ln \overline{W}_n - \ln \overline{W}_o = (\overline{X}_n - \overline{X}_o)\hat{\beta}_o + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}_n + (\hat{\beta}_{\lambda n} \overline{\lambda}_n - \hat{\beta}_{\lambda o} \overline{\lambda}_o) \qquad [5.16]$$

The first term on the right hand side of equation [5.16] [$(\overline{X}_n - \overline{X}_o)\hat{\beta}_o$] is the contribution to the difference in wages that can be explained by the mean differences in characteristics between nurses and other workers (the difference in the X's). This is referred to as the difference due to endowments, or the difference in variables. In equation [5.16] these characteristics are evaluated using the co-efficients of all other workers. The second term in equation [5.16] [$(\hat{\beta}_n - \hat{\beta}_o)\overline{X}_n$] provides a measure of the premium to being a nurse. That is, this gives the actual mean wage of nurses with mean characteristics and what they are predicted to earn if the characteristics are weighted by the returns of other workers. It is the contribution to the differences in wages that can be explained by the differences in returns to characteristics (the β's). Note that this premium also captures the difference in intercept terms between the two wage equations $(\hat{\beta}_{n0} - \hat{\beta}_{o0})$. The third term in equation [5.16] $[(\hat{\beta}_{\lambda n}\overline{\lambda}_n - \hat{\beta}_{\lambda o}\overline{\lambda}_o)]$ describes the effect of potential selection bias on wage differentials between nurses and other workers. We shall call this the differences due to selection bias.

This decomposition does not provide a unique disaggregation of the difference in wages between nurses and other workers. Alternatively, the wage decomposition described in equation [5.16] may be reversed and the premium can be computed by comparing what non-nurses would earn if their characteristics were weighted by the returns of nurses. This yields the following decomposition:

$$\ln \overline{W}_n - \ln \overline{W}_o = (\overline{X}_n - \overline{X}_o)\hat{\beta}_n + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}_o + (\hat{\beta}_{\lambda n}\overline{\lambda}_n - \hat{\beta}_{\lambda o}\overline{\lambda}_o) \qquad [5.17]$$

where the differences due to endowments, the premium, and the differences due to selection bias are described analogously. In general the two decomposition methods will yield different estimates of the premium to being employed as a nurse. This can be seen more clearly with reference to Figure 5.1 which is constructed under the assumption that nurses' earnings are greater than the earnings of other workers ($\ln \overline{W}_n > \ln \overline{W}_o$). From Figure 5.1 we have the following decompositions:

[14] As noted in Chapter 4 this is why Heckman's two-step approach is adopted here rather than the FIML approach.

$$P_1 = [(\hat{\beta}_{n0} - \hat{\beta}_{o0}) + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}_n] \quad [5.18]$$

$$P_2 = [(\hat{\beta}_{n0} - \hat{\beta}_{o0}) + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}_o] \quad [5.19]$$

$$E_1 = (\overline{X}_n - \overline{X}_o)\hat{\beta}_o \quad [5.20]$$

$$E_2 = (\overline{X}_n - \overline{X}_o)\hat{\beta}_n \quad [5.21]$$

For simplicity the differences due to selection bias are omitted. The premium to being employed as a nurse $[(\hat{\beta}_{n0} - \hat{\beta}_{o0}) + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}]$ will differ according to whether we evaluate the differences in returns to characteristics in terms of nurses' (P_1) or other workers' (P_2) characteristics. Note that if the returns to given characteristics are the same $(\hat{\beta}_n = \hat{\beta}_o)$ then the premium to being employed as a nurse is measured by the difference in intercept terms $(\hat{\beta}_{n0} - \hat{\beta}_{o0})$ and the source of this is unknown.

To summarise, the difference in earnings between nurses and other workers is due to:

1. The differences in measured productive characteristics $(\overline{X}_n - \overline{X}_o)$;
2. The difference in intercept terms $(\hat{\beta}_{n0} - \hat{\beta}_{o0})$; and/or,
3. The difference in returns that nurses receive for any given characteristics $(\hat{\beta}_n - \hat{\beta}_o)$.

The decomposition analysis allows us to disaggregate any earnings differential into these different components. [15]

[15] Note the relevance to an analysis of nurses' relative earnings. An accurate comparison of relative earnings relies on the selection of an appropriate comparator. We compare in the analysis the earnings of nurses and the earnings of all other workers combined. The comparator is unlikely to be appropriate because nurses might earn higher mean wages but this is because they have superior individual and labour market endowments. The problem is that we are not comparing like with like. The decomposition analysis compensates for this effect and controls for the impact of differences in individual and labour market characteristics on earnings. By examining

The two decompositions (equations [5.16] and [5.17]) will yield identical estimates of the premium to being employed as a nurse only if the following condition holds:

$$(\overline{X}_n - \overline{X}_o)(\hat{\beta}_n - \hat{\beta}_o) = 0 \qquad [5.22]$$

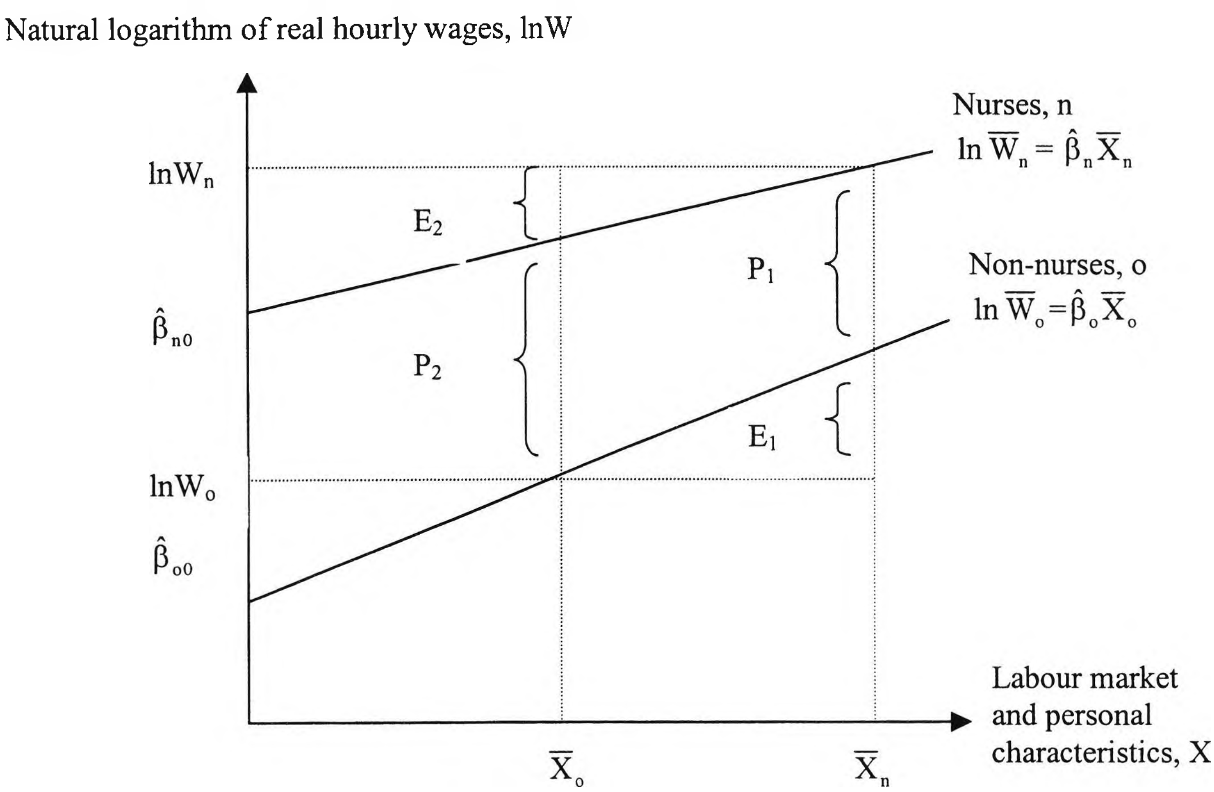


Figure 5.1. A graphical representation of the decomposition analysis

the 'premium' component of the decomposition we ascertain whether nurses are paid comparable wages after controlling for differences in individual and labour market characteristics. We are able to compare in terms of their earnings nurses' with the 'same' workers in other occupations. Thus intuitively we are now comparing like with like.

Equation [5.22] is satisfied only if the returns to characteristics (the $\hat{\beta}$'s, including the intercept terms) are identical for nurses and other workers, or if the mean characteristics themselves (the $\overline{X}$'s) are identical. It is possible to test whether the $\overline{X}$'s are the same using t-tests on the differences in sample means (see section 5.6). It is possible to test whether the $\hat{\beta}$'s are the same using the Chow test. This test allows us to ascertain whether the relationship between lnW and X is different for nurses and all other workers. It also allows to ascertain whether if there is a difference in the $\hat{\beta}$'s it is in the intercepts, the slopes, or both. In section 5.8 we conduct Chow tests for structural differences between the wage equations for nurses and all other workers.

We decompose the wage differential between nurses and other workers via both methods using the estimates generated from Models 3-5. For Model 3 we omit the third term on the right hand side of equations [5.16] and [5.17].

5.4. The data

The data used to estimate Models 1-5 were taken from the Quarterly Labour Force Survey (QLFS) from Winter 1992 to Autumn 2000. [16] The Labour Force Survey (LFS) is a survey of households living at private addresses in Great Britain. Its purpose is to provide information on the UK labour market that can then be used to develop, manage, evaluate and report on labour market policies. The LFS is conducted by the Social Survey Division of the Office for National Statistics (ONS) in Great Britain and by the Central Survey Unit of the Department

[16] The author did originally conduct the analysis using the British Household Panel Survey (BHPS), waves 1 to 7. The data were obtained and formatted and each statistical model was constructed and applied. Unfortunately due to the size of the survey the number of nurses in the BHPS sample was small (n = 110) and the results were therefore disappointing. Hence the QLFS – a much larger survey with many more nurses – was used instead.

of Finance and Personnel in Northern Ireland. The first LFS was conducted in 1973 and then every two years to obtain information which could assist in the framing and monitoring of social and economic policy. By 1983 it was being used by the Department of Employment to obtain measures of unemployment on a different basis from the monthly claimant count and to obtain information that was not available from other sources or was only available for census years, for example, estimates of the number of people who were self-employed. Between 1984 and 1991 the survey was carried out annually and consisted of two elements: (1) a quarterly survey of approximately 15,000 private households conducted in Great Britain throughout the year; and, (2) A 'boost' survey in the quarter between March and May, of over 44,000 private households in Great Britain and 5,200 households in Northern Ireland. Quarterly compilation of LFS estimates for Great Britain (that is, the QLFS) became possible in 1992 when the sample was increased to cover 60,000 households living in Great Britain every quarter.

The LFS is representative of the population of the UK. The population covered is all individuals resident in private households, all individuals resident in National Health Service accommodation and young people living away from the parental home in a student hall of residence or similar institution during term time. The sample design currently consists of approximately 59,000 randomly selected households in Great Britain every quarter, representing 0.3% of the population. A sample of approximately 2,000 responding households in Northern Ireland is added to this (representing 0.4% of this population), allowing United Kingdom analyses to be made.

The QLFS data refer to the seasonal quarters March-May (Spring), June-August (Summer), September-November (Autumn) and December-February (Winter). Each QLFS sample of

61,000 households is made up of five 'waves', each of approximately 12,000 private households. Each wave is interviewed in five successive quarters, so that in any one quarter, one wave will be receiving their first interview, one wave their second, and so on, with one wave receiving their fifth and final interview. Thus there is an 80% overlap in the samples for each successive quarter.

Households are interviewed face to face at their first inclusion in the survey and by telephone at quarterly intervals thereafter. Households have their fifth and final quarterly interview on the anniversary of the first.

Questions asked in the QLFS cover the following broad topics:

- Individual characteristics (e.g. sex of respondent, age, marital status, nationality, country of birth, ethnicity, place of residence);
- Household and family characteristics (e.g. type of household, composition of household, number of persons in household, number of dependent children in household, housing tenure);
- Economic activity (e.g. employment status);
- Characteristics of main job (e.g. industry, occupation, public or private sector);
- Education and training (e.g. qualifications, highest qualification, age completed full-time education);
- Health (e.g. health problems, whether health problems limit activity); and,
- Income (e.g. gross and net income, benefits, occupational pension, other income).

In terms of formatting the QLFS data for the present analysis the following issues were relevant:

1. Income/wage questions were included in the QLFS only from Winter 1992/93 onwards. Since wages are the key dependent variable in this analysis only the QLFS data from Winter 1992/93 and following are used.
2. Income/wage questions in the QLFS are only asked of respondents receiving their fifth and final interviews because of concerns that the questions might have an adverse impact on participation in the survey and overall response rates. Since wages are the key dependent variable in this analysis only individuals for whom wage data are available (or would be available if they worked) are included. This means that only individuals in their fifth and final wave are included. The analysis therefore utilises approximately one fifth of the total QLFS sample.
3. Income/wage questions in the QLFS are asked only of employed individuals. Importantly, the self-employed are not asked about their earnings due to concerns of the effect this might have on participation in the QLFS in this sub-group of working individuals. This means that a proportion of individuals who do participate in the labour market were excluded from the analysis due to a lack of data on the key dependent variable (wages).
4. It is important to note that even though individuals are included in the QLFS for five successive waves, because the sample of data used in the present analysis consists only of individuals in the fifth and final wave, these data are not a panel (a time-series cross-section). The data used do not consist of multiple observations across time on each of many cross-sectional observational units (individuals). Instead they consist only of single observations at one (albeit different) point in time. This means that we are unable to control for individual heterogeneity in the data through the use of, say, random or fixed

effects, though clearly some method is required of dealing with the heterogenous nature of the time dimension in the data.

5. A number of errors have been found by the ONS in the QLFS data in the responses to some of the household and family characteristics questions that led to a number of variables being deleted from the dataset at source. Most importantly in terms of the present analysis information on the number of dependent children has been removed from the original QLFS data. This is unfortunate since having to look after dependent children may well affect the reservation wage of (particularly) females and is therefore likely to be an important factor explaining female participation in the labour market. A proxy variable is included instead to account for this factor (see below).

6. As explained in Chapter 4 property income N is an important component in the determination of participation in the labour market. For individuals this is measured directly in the QLFS in terms of receiving an occupational pension and receiving non-labour or unearned income from any other source. In modelling female labour supply, one frequently considered framework within the group of 'family models' is the so-called chauvinistic model of labour supply. In this model it is assumed that one of the partners (typically the male) maximises their utility independently of the choices of the other partner. The other partner (typically the female) then maximises their utility $U=U(C, L)$ subject to a budget constraint $WH+N=C$ which now includes their partner's earnings (N_P) in addition to their own non-labour income (N_O) as property income. That is, $N=N_P+N_O$. If we accept the chauvinistic model as correct this means that partners' labour income should be included as an independent variable (or at least included in the calculation of the property income variable) in the female participation equation. Unfortunately in the QLFS it is not possible to discern partners' earnings. Of course, this omission in the data is not problematic at all if we reject the chauvinistic model. Indeed,

this model has been criticised because it has little to say about changing attitudes and culture which have tended to increase in recent years the participation of women, effectively undermining the realism of the model (Bosworth et al., 1996). In this case an alternative model – such as the individual utility maximisation model – may be more appropriate.

7. As explained above we include in the final sample only working individuals in their fifth and final wave. We must also include non-working individuals in their fifth and final wave in order to estimate the probit models of participation required for the participation selection bias models. To include non-working individuals in other waves would mean that individuals would appear more than once in the final sample. Unfortunately, in the QLFS dataset information is missing on which wave responding individuals are in for the period Autumn 1993 to Winter 1996/97. Workers with observed wages are clearly identified as being in the fifth and final wave. However, for non-workers it is not possible to determine which wave an individual belongs to. To solve this problem we include all non-working individuals from Autumn 1993, Winter 1994/95 and Spring 1996 (quarters 3, 8 and 13) and no non-working individuals for the periods Winter 1993/94-Autumn 1994, Spring 1995-Winter 1995/96 and Summer 1996-Winter 1996/97 (quarters 4, 5, 6, 7, 9, 10, 11, 12,14, 15 and 16). The assumption is that, for example, non-working individuals in wave 4 in Autumn 1993 are the same as non-working individuals in wave 5 in Winter 1993/94, that non-working individuals in wave 3 in Autumn 1993 are the same as non-working individuals in wave 5 in Spring 1994, and so on. In terms of the statistical analyses what this means is that the quarterly time trend dummy variables (see below) for the quarters 4, 5, 6, 7, 9, 10, 11, 12,14, 15 and 16 predict perfectly participation in the labour market (because all individuals in these quarters participate). These variables are therefore dropped from the participation

equations.

8. For estimation of Model 4 it is necessary to estimate separate probit models of participation for nurses and all other workers. This means that it is necessary to have information on non-participating nurses (and also non-participating non-nurses). Unfortunately in the QLFS dataset respondents are only asked for details of their occupation if they are participating in the labour market at the time of the survey. It is therefore not straightforward to define non-participating nurses. For the purposes of the analysis we therefore adopt a pragmatic approach and define non-participating nurses as individuals who have a nursing qualification but who are not participating in the labour market. Non-participating non-nurses are then defined as all other non-participators.

The final sample in the QLFS data comprises females aged 18 to 60 years. Females only are selected since in the British nursing labour market a high proportion of nurses (around 90%) are females. Females below the age of 18 and above the age of 60 are excluded since labour market participation rates of females in these age groups are very low (ONS, selected years).

5.5. The variables

5.5.1. The wage equations

The key variables included in the wage equations, according to the Mincerian earnings function model described in Chapter 4 are: wages; years of education and, labour market experience. Additionally, the relevant occupational choice variable (i.e. being employed as a nurse or not) is included.

The wage used (LNWAGE) is the natural logarithm of the hourly wage measured in constant December 1992 UK£. Hourly wages were computed as usual gross weekly pay divided by total usual hours worked per week. Hourly wages were used (instead of weekly, monthly or annual wages) to allow for the effect of total hours worked on total wages. Wages were converted to constant December 1992 prices using the monthly retail prices index (Office for National Statistics, selected years). Years of education (YED) were computed by subtracting school starting age (assumed to be age 4 years) from age completed full-time education. Also included is a quadratic term (years of education squared, YED2) to allow for a possible concavity in the earnings-years of education profile. This is used in preference to other functional forms (e.g. higher order polynomials, piecewise linear regression) because it is grounded in the Mincerian economic model of earnings, as discussed in Chapter 4 and Appendix 4.1, which has proved many times to be consistently useful as an empirical tool for the analysis of the determinants of wages. The other key variable in the Mincerian earnings function is work experience. This may be measured/proxied in a number of ways, including: age; age minus years of schooling minus six; years since leaving school; years working in a specific occupation; years working with a specific employer; and, years working in a specific job. Each of these has some merit as a measure of labour market experience. Age is the simplest and most easily obtainable measure but clearly overestimates post-full-time education experience. Both age minus years of schooling minus six and years since leaving school give a more accurate depiction of work experience, but fail to allow for non-participation in the labour force and for employment in occupations unrelated to the current occupation where such experience may not be relevant to earnings. Years working in a specific occupation or with a specific employer or in a specific job are important if experience specific to that occupation/employer/job is relevant to current earnings. Where experience outside of these is relevant these measures will underestimate the true level of experience. In a cross-section of the population, each of these measures are likely to be most appropriate for some proportion of workers. To this extent, all the above measures of work experience are applicable. In this analysis years of experience (EXP) were computed as the number of years employed with the current employer, estimated as the 'current' year minus year started working with the current employer. An additional variable (years of experience squared, EXP2) is included in the wage equation to allow for possible concavities in the experience-earnings profile.

The other key variable included in the wage equations for Models 1-2 is a dummy variable capturing whether or not the individual is employed as a nurse in the NHS. NURSE was computed using the Standard Occupational Classification (SOC) code of the occupation in main job of all workers in the sample. SOC code 340 ('registered nurse') was used to define qualified nurses in the analysis. Unqualified nurses (nursing auxiliaries and assistants) were not included since their training, qualifications, job specification, work-related skills, and pay are significantly different to that of qualified nurses. Also excluded from the sub-sample of nurses in the data were a small number of private sector nurses (private sector nurses constitute 15% of all nurses). Because they work in a separate labour market to NHS nurses with different job specifications and different job characteristics (for example, 85% of private sector nurses work in nursing homes) for the purposes of the analysis they were counted as 'non-nurses'.

Described above are the key variables included in the wage equation as determined by economic theory and the empirical question to hand. Other control variables are included in the wage equation that may have an impact on wages. The inclusion of these variables is informed by the review of the empirical literature on extended earnings functions presented in the previous chapter. The additional variables included are educational attainment variables (NURSEQUA, PGDEG, DEG, ALEVEL and NOQUAL), personal characteristic variables (DISABLE, ETHNIC, NONBRIT and ETHNBRIT), a regional dummy variable for whether or not the individual lives in the South East of England (SEAST), and job characteristic variables (HOURSPW, MANAGE, NWORKERS and TEMP). The definition of each of these control variables is given below. NURSEQUA, a dummy variable for whether or not the individual has a nursing qualification, is included in the wage equations for Models 3-5 only, which estimate separate wage equations for nurses and workers in other occupations. This variable is included in these models because one might expect the earnings of nurses to be affected by whether or not they have a nursing qualification (having a nursing qualification is likely to be a basic requirement for any place on the nursing promotional ladder). This variable is not included in the wage equations for Models 1-2 because this might dilute the effect of being employed as a nurse (captured by the NURSE dummy variable), which is the factor of interest in the present analysis.

An additional set of control variables (time trend variables) is included to allow for the heterogenous nature of the time dimension in the data. As explained above the data consists of independent cross-sections comprising single observations on different cross-sectional units (individuals) at different points in time. Some method is warranted of allowing for a possible time trend in the data with respect to wages. A set of dummy variables is included in the wage equations specifying the quarter to which the individual data pertain. The QLFS data are pooled across 32 quarters between Winter 1992 and Autumn 2000. 31 dummy variables in total are included for each consecutive quarter (Q1, Q2, Q3,..., Q31). To avoid perfect collinearity via the dummy variable trap a dummy variable for the first quarter (Q0) is omitted. The time periods to which each quarter pertains are given in Appendix 5.1.

5.5.2. The participation equations

The key variables to be included in the participation equation, according to the model of individual labour supply described in Chapter 4 are: participation in the labour market; wages; and, property income.

Participation in the labour market (PART) is measured by a dummy variable reflecting whether or not an individual is an employee. As explained above, the self-employed are excluded from the sample because wage data are not collected on such individuals in the QLFS. Reasons for non-participation include: because the individual is unemployed; because the individual is an unpaid family worker (e.g. housewife); or, because the individual is unavailable for or not seeking work because they are sick, disabled or looking after their family.

It is not appropriate to include wages in the participation equations estimated here since a summary measure of participation [$\lambda(p)$] is then inserted into the wage equation. Inclusion of wages in the participation equation would mean that the same variable is then effectively included as both an independent variable and the dependent variable in the same wage equation.

Two property income variables are included in the participation equation: a dummy variable for whether or not the individual receives an occupational pension (PENSION); and another variable measuring the amount of other non-labour income received in the last 12 months (NONLABY). NONLABY includes unearned income accruing from the ownership of assets and from state payments (such as Child Benefit in the early 1990s) that are payable regardless of the number of hours worked and/or income received. This variable also includes elements such as income from shares and from property rent. It does not include state payments the magnitude of which depends on the number of hours worked and/or the income received (such as income support or the job seeker's allowance). Note that there is likely to be a positive relationship between age and whether or not the individual receives a pension.

A number of other variables are included in the participation equation. Individual participation is determined by comparisons between the offered market wage and the individual's reservation wage, given individual preferences and the level of property income. (see Chapter 4). The reservation wage, since it cannot be observed directly in the data is proxied by variables that are likely to affect its value. These include age variables (AGE, AGE2), personal characteristic variables (DISABLE, ETHNIC, NONBRIT and ETHNBRIT), family variables (PCHILD, COHABIT, MARRIED), education variables (YED, YED2, PGDEG, DEG, ALEVEL and NOQUAL), and a dummy variable for whether or not the individual lives in the South East of England (SEAST). The definition of each of these variables is given below. Additionally a set of 31 dummy variables (Q1, Q2, Q3,..., Q31) are included specifying the quarter to which the individual data pertain between Winter 1992 and Autumn 2000 to allow for a possible time trend in the data with respect to participation.

Since information on the numbers of dependent children has been removed from the original QLFS data (see section 5.4) an alternative proxy measure is introduced to capture the affect that having children has on participation through the costs they impose on employment and via their influence on preferences. PCHILD is a dummy variable taking the value one if the individual female is aged 20 to 29 years and cohabiting or aged 25 to 34 years and married and zero otherwise. This specification was chosen because in excess of 60% of all births occur to mothers in these circumstances (ONS, 2001). It is unfortunate that information on the number of dependent children is not directly available in the original QLFS data. Nevertheless we adopt a pragmatic approach in an attempt to capture this effect.

5.5.3. The occupation selection equation

The key variables to be included in the occupation equation, according to the occupation selection bias model described in Chapter 4 are: individual productive characteristics affecting wages; personal characteristics that affect occupational choice; individual factors affecting tastes for each occupation and, ability relevant to being employed as a nurse and in some other occupation.

Individual productive characteristics are measured via education and training. NURSEQUA is a dummy variable for whether or not the individual has a nursing qualification. Other measures of education that are included are years of education variables (YED, YED2) and educational attainment variables (PGDEG, DEG, ALEVEL and NOQUAL).

Other variables are included in the occupation selection equation that may have an impact on the decision to become a nurse, either as personal characteristics affecting occupational choice or as individual factors affecting tastes for each occupation. These are age variables (AGE, AGE2), personal characteristic variables (DISABLE, ETHNIC, NONBRIT and ETHNBRIT), family variables (PCHILD, COHABIT, MARRIED), property income variables (PENSION, NONLABY), a dummy variable for whether or not the individual lives in the South East of England (SEAST), and a set of 31 dummy variables (Q1, Q2, Q3,..., Q31) specifying the quarter to which the individual data pertain between Winter 1992 and Autumn 2000 to allow for a possible time trend in the data with respect to occupation selection. The definition of each of these variables is given below.

Ability is not measured directly in the data. Instead we include an additional regressor [λ(nu)] that corrects the market wage in a specific occupation to account for an individual's propensity to be employed in that occupation. As discussed in Chapter 4 this formulation is suggested by the theory of comparative advantage in terms of the ability relevant to being employed as a nurse or in some other occupation.

5.5.4. Identification of the wage equations

In Chapter 4 Section 4.12.1 it was noted that a lack of exclusion restrictions in the Heckman two-step model is likely to lead to identification problems with the wage equations. Exclusion restrictions identify the two-step model, where in the participation equations and the occupation selection equations variables are included that are excluded from the wage equations. Appropriate identifying variables will influence the individual's participation and occupational choice decision without influencing earnings. We follow the labour supply literature and base identification of the wage equation on exclusion of the property income variables (PENSION and NONLABY). A priori it is posited that these variables will be good (i.e. statistically significant) predictors of participation and occupation selection but will not be associated with actual observed wages (the dependent variable in the wage equation). In the context of the participation selection bias models underpinning this is the idea that these factors will affect labour force participation via the reservation wage rather than the offered wage. In the context of the occupation selection bias model the a priori belief is that the property income variables will affect attitudes to specific occupations and job characteristics and will therefore be important explanatory variables in occupation selection but will not affect wages directly. The assumption is that property income enters the utility function either via tastes for each occupation or as an argument of the personal characteristics that are not also included in the individual productive characteristics. This a testable hypothesis in the data. For each of the three models that utilise Heckman's two-step approach (Models 2, 4 and 5) we re-run the wage equation with no exclusion restrictions (see below).

5.6. Descriptive statistics

The definitions and descriptive statistics of the variables used in the participation equations for Models 2 and 4 and the occupation selection equation for Model 5 are presented in Appendix 5.2. The descriptive statistics used for the wage equations are presented in Table 5.1. The entire sample of all workers and non-workers used in the analysis consist of 247,774 females aged 18 to 60 years. Of these, 61% (151,944) participate in the labour market. Of the 151,944 workers 4% (6,608) are employed as NHS nurses. Using the definition of non-participating nurses given above as individuals with a nursing qualification who do not work there are 8,878 nurses in the data of whom 2,270 (26%) do not participate. This participation rate is higher than for the 238,896 non-nurses, of whom 61% (145,336) participate.

Table 5.1 presents the main variables of interest in the present analysis – the descriptive statistics of variables used in the wage equations. Nurses are paid on average higher wages than workers in all other occupations combined. [17] In the sample the mean real hourly wages of nurses and all other workers are £7.36 (Std. Dev. £2.96) and £5.49 (Std. Dev. £3.50), respectively (data not shown). The difference in mean real hourly wages (£1.87 – nurses receive on average 34% higher wages than all other workers) is statistically significant at conventional levels ($p<0.0001$, 95% confidence interval £1.80 to £1.95). In the present analysis the dependent variable in the wage equations is the natural logarithm of the real hourly wage (LNWAGE). As presented in Table 5.1 the mean value of LNWAGE for nurses in the sample is 1.9320 and for all other workers combined it is 1.5672. This difference in mean LNWAGE (0.3648) is statistically significant at standard levels ($p<0.0001$; 95% confidence interval 0.3552 to 0.3745). The main focus of the decomposition analysis is to examine the extent to which the different characteristics of nurses and other workers can explain this wage differential. It is therefore informative to compare initially the labour market and personal characteristics of nurses versus all other workers, as used in the wage equations.

Most of the differences between nurses and all other workers in the means of the variables included in the wage equations are statistically significant. In terms of education nurses and other workers have comparable years of full-time education (mean approximately 13 years). A greater proportion of nurses has a nursing qualification than the rest of the working population. Nurses are generally less well educated at the top end of the educational attainment spectrum in the sense that a lower proportion of nurses have postgraduate degrees (1% versus 3%), first degrees (6% versus 9%) and A levels (1% versus 7%) as their highest educational qualification. However, at the other end of the spectrum nurses are better educated than other workers because a greater proportion of nurses do possess some form of educational qualification (99% versus 84%).

---

[17] In the decomposition analysis we will determine whether this is due to nurses' superior human capital

According to the economic model of earnings described above another important factor likely to influence earnings is years of work experience. Nurses on average tend to have more years of work experience than other workers (mean 12 years versus mean 10 years).

In terms of the personal characteristics variables nurses and other workers are comparable in terms of the prevalence of health problems affecting paid work of workers (approximately 5%). Slightly more nurses are from non-white ethnic groups (5% versus 3%) and have a non-British nationality (6% versus 3%). Slightly fewer nurses live in the South East of England (26% versus 30%) which might be important because individuals living in this geographical area receive extra wage payments in terms of a London weighting allowance to help cover the increased cost of living in this region.

Nurses tend to work longer hours than other workers (mean 33 versus mean 31 total hours per week), and a larger proportion of nurses play some kind supervisory role in their job (76% of nurses are employed as a supervisor, manager or foreman compared to 24% of all other workers). Nurses on average tend to work in larger establishments in terms of numbers of workers employed (83% of nurses work in establishments with 25 or more total workers, compared with 62% of other workers), and a slightly smaller proportion of nurses are employed on temporary contracts (6% versus 7%). In terms of the time trend variables (not shown), each quarter in the dataset contains between 2% and 5% of all observations (workers). The proportions are similar across nurses and all other workers (that is, the difference in proportions are not statistically significant in the majority of cases). The full set of descriptive statistics of variables used in the wage equations, including the time trend variables, are presented in Appendix 5.2.

---

endowments or whether there is a premium to being employed as a nurse.

	All workers [1]		Nurses only [2]		Other workers only [2]		Definition
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	
LNWAGE*	1.5830	0.5138	1.9320	0.3835	1.5672	0.5133	LN hourly wage
NURSE	0.0435	0.2040					Employed as a nurse=1, 0 otherwise
Years of education variables							
YED*	13.2908	2.5012	13.5051	1.9672	13.2811	2.5224	Years of full-time education
YED2*	182.9020	75.1721	186.2580	59.3591	182.7490	75.8093	Years of full-time education squared
Educational attainment variables							
NURSEQUA*			0.9215	0.2690	0.0243	0.1539	Has a nursing qualification=1, 0 otherwise
PGDEG*	0.0315	0.1747	0.0135	0.1153	0.0323	0.1768	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG*	0.0922	0.2894	0.0642	0.2451	0.0935	0.2912	Highest qualification is a first degree=1, 0 otherwise
ALEVEL*	0.0685	0.2526	0.0101	0.1002	0.0712	0.2571	Highest qualification is A level=1, 0 otherwise
NOQUAL*	0.1540	0.3609	0.0045	0.0672	0.1608	0.3673	Has no qualifications=1, 0 otherwise
Work experience variables							
EXP*	9.7689	7.0677	11.9953	7.9082	9.6677	7.0103	Years of experience with current employer
EXP2*	145.3830	210.2600	206.4170	253.3180	142.6080	207.6660	Years of experience with current employer squared
Personal characteristics variables							
DISABLE#	0.0525	0.2230	0.0478	0.2134	0.0527	0.2234	Health problems affect paid work =1, 0 otherwise
ETHNIC*	0.0359	0.1860	0.0528	0.2237	0.0351	0.1841	Non-white ethnic group=1, 0 otherwise
NONBRIT*	0.0366	0.1878	0.0605	0.2385	0.0355	0.1851	Non-British nationality=1, 0 otherwise
ETHNBRIT*	0.0093	0.0958	0.0182	0.1335	0.0089	0.0937	Non-white and non-British=1, 0 otherwise
Regional variables							
SEAST*	0.3031	0.4596	0.2624	0.4400	0.3049	0.4604	Lives in the South East of England=1, 0 otherwise
Job characteristic variables							
HOURSPW*	31.2637	12.9985	33.5796	10.5943	31.1584	13.0876	Total usual hours worked per week
MANAGE*	0.2696	0.4438	0.7639	0.4247	0.2472	0.4314	Employed as a supervisor, manager or foreman=1, 0 otherwise
NWORKERS*	0.6347	0.4815	0.8390	0.3676	0.6254	0.4840	25+ workers at workplace=1, 0 otherwise
TEMP*	0.0756	0.2643	0.0610	0.2393	0.0762	0.2654	Job is non-permanent or temporary=1, 0 otherwise
N	151,944		6,608		145,336		

[1] Wage data for all workers are used in Models 1-2

[2] Separate wage data for nurses only and for all other workers only are used in Models 3-5

* Difference in mean values between nurses and all other workers significant at the 5% level

\# Difference in mean values between nurses and all other workers significant at the 10% level

Table 5.1. Descriptive statistics of variables in wage equations

5.7. Results of statistical models

To aid the exposition we include here a summary of the main features of each model reported in this section.

Model	Structure	Sample	Estimation
1	Wage equation with dummy variable for whether or not an individual is employed as a nurse	Workers only	OLS
2	Participation equation and wage equation with dummy variable for whether or not an individual is employed as a nurse	Workers and non-workers (participation equation), workers only (wage equation)	Heckman two-step procedure
3	Separate wage equations for nurses and all other workers	Workers only	OLS
4	Separate participation equations and wage equations for nurses and all other workers	Workers and non-workers (participation equation), workers only (wage equations)	Heckman two-step procedure
5	Occupation selection equation, and separate wage equations for nurses and all other workers	Workers only	Heckman two-step procedure

Table 5.2. Main features of Models 1-5

For each set of results the co-efficients for the time trend variables are not shown. See Appendix 5.3 for the full set of results including those pertaining to the time trend variables.

5.7.1. Model 1

Table 5.3 reports the results of the first statistical model. This is a simple OLS earnings function estimated across all workers in the sample including a dummy variable for whether or not an individual is employed as a nurse. The explained variation in ln wages is approximately 30%, which is comparable with similar studies (Shields and Wheatley Price, 1998). The OLS results are corrected for potential heteroscedasticity using White's robust covariance matrix (White, 1980).[18] All variables except for some of the time trend variables (not shown – see Appendix 5.3) are statistically significant at the 5% level. The co-efficients are of the expected sign and order of magnitude and reveal similar patterns to those in other UK earnings function studies in the literature.

NURSE is included to estimate the wage premium to being employed as a nurse at the sample mean of all workers, controlling for measurable individual productive characteristics. Since the model is in the semi-log form and NURSE is a dummy variable the co-efficient on NURSE (0.2034) needs to be transformed to obtain the percentage change in earnings (see Appendix 5.4). In this case individuals employed as nurses earn on average 23% (= $[e^{0.2034} - 1]*100$) higher wages than those employed in other occupations. This suggests that the premium to being employed as a nurse is 23%, controlling for other measurable individual productive characteristics (but not adjusting for selection bias). Note that this premium is lower than the 34% difference in unadjusted wages between nurses and all other workers described above.

Other co-efficients in Model 1 are consistent with the general Mincerian model of earnings. First, the signs and values of these co-efficients on the years of education variables indicate that the relationship between years of education and earnings is non-linear, and that the earnings-years of education profile is in fact n-shaped with maximum earnings occurring following 19 years of education ($= \frac{-0.1367}{2*-0.0036}$, see Appendix 5.5). Second, the co-efficients on the work experience variables also indicate a concavity in the experience-earnings profile. In this case the relationship between work experience and earnings is also n-shaped with earnings maximised at 33 years of work experience.

[18] This approach is advocated by Greene (2000). The correction is also applied to Model 3.

The educational attainment variables are also important factors influencing earnings. Returns to educational attainment by type are of the expected rank order. Across all workers the transformed co-efficients suggest that the wage premium to obtaining a postgraduate degree is 47% and that for a first degree is 34%. Having A levels as the highest educational qualification leads to a small positive wage premium (2%). Workers with no qualifications earn on average 13% lower wages than their better-educated counterparts.

The co-efficient on DISABLE in the personal characteristic variables indicates that working individuals with health problems that affect paid work will on average earn 8% lower wages than non-disabled workers. Non-white British workers earn slightly lower (-1%) wages than white British workers and white non-British workers earn slightly higher wages (5%). The co-efficient on ETHNBRIT indicates that the mean earnings of non-white non-British workers is different from the mean earnings of other non-white or non-British workers. For example, non-white workers earn slightly less than their white counterparts, but their earnings are even lower (-10%) if they also happen to be non-British.

From the regional variables we can see that workers living in the South East of England earn on average 17% higher wages than those living outside of this area. This will partly reflect additional wage payments in terms of a London weighting allowance to help cover the increased cost of living in this region.

In terms of the job characteristic variables, the co-efficient on HOURSPW indicates that on average across all workers, hourly wages increase with the number of hours worked. There is a wage premium to being employed in a managerial position (supervisors, managers or foremen earn on average 15% higher wages than their more junior counterparts), and the positive co-efficient on NWORKERS indicates that the larger the workplace the larger are average earnings. Workers on temporary contracts receive slightly higher earnings (by 1%) than those on permanent contracts. The time trend variables (not shown – see Appendix 5.3) are of the expected sign (positive) relative to the base quarter (quarter 0, Winter 1992/3) and are generally of the expected rank order of magnitude. The co-efficients on these variables indicate that real wages are generally, but by no means universally or uniformly, increasing over time.

	β [1]	Std. Err. [2]
Constant	-0.1255*	0.0458
NURSE	0.2034*	0.0052
Years of education variables		
YED	0.1367*	0.0061
YED2	-0.0036*	0.0002
Educational attainment variables		
PGDEG	0.3900*	0.0077
DEG	0.2904*	0.0051
ALEVEL	0.0164*	0.0052
NOQUAL	-0.1427*	0.0033
Work experience variables		
EXP	0.0326*	0.0005
EXP2	-0.0005*	0.0000
Personal characteristic variables		
DISABLE	-0.0802*	0.0054
ETHNIC	-0.0151*	0.0073
NONBRIT	0.0507*	0.0079
ETHNBRIT	-0.1007*	0.0163
Regional variables		
SEAST	0.1580*	0.0026
Job characteristic variables		
HOURSPW	0.0008*	0.0001
MANAGE	0.1383*	0.0028
NWORKERS	0.1355*	0.0025
TEMP	0.0142*	0.0052
Adjusted R^2	0.2985	
Model test	$F(49, 151{,}894) = 1320.57$; $p = 0.0000$	
N	151,944	

[1] Dependent variable is LNWAGE
[2] Results corrected for heteroscedasticity using White's estimator
* Significant at the 5% level

Table 5.3. Results of Model 1: OLS estimates of wage equation [5.1] based on all workers with NURSE dummy [3]

5.7.2. Model 2

Model 2 is estimated using the participation selection bias model including a dummy variable in the wage equation for whether or not an individual is employed as a nurse. The results of the participation equation are presented in Table 5.4. In terms of the structural model of individual labour supply the key variables in the participation equation are the property income variables. The co-efficients on these variables are statistically significant and of the expected sign: holding all other variables constant, first, if an individual receives an occupational pension, or second, as non-labour income increases, the less likely the individual is to participate in the labour market.

Examining the co-efficients on the other variables included in the participation equation, all co-efficients except those on ETHNBRIT and SEAST are statistically significant. Holding all other variables constant, the co-efficients on the age variables (AGE and AGE2) indicate there is a concavity between an individual's age and their propensity to participate in the labour market. From the personal characteristic variables, the disabled, non-whites and non-British individuals are less likely to participate in the labour market. The non-significance of ETHNBRIT indicates there is no additional interaction effect on participation of being both non-white and non-British. In terms of the family variables, the negative co-efficient on PCHILD supports the prior expectation that having children reduces the likelihood of labour market participation. The co-efficients on COHABIT and MARRIED are positive, indicating that cohabiting and married women are more likely to participate than single women are. This is perhaps surprising, but the effect of these factors may be interpreted slightly differently if it is also borne in mind that PCHILD is basically an interaction term capturing the joint impact of marital/cohabiting status and age combined (essentially PCHILD picks up the effect of being aged 20 to 29 years and cohabiting or aged 25 to 34 years and married. COHABIT therefore picks up the effect of cohabiting outside the age range 20-29 years and MARRIED picks up the effect of being married outside the age range 25-34 years).

	δ [1]	Std. Err.
Constant	-1.1742*	0.0721
Age variables		
AGE	0.0748*	0.0020
AGE2	-0.0008*	0.0000
Personal characteristic variables		
DISABLE	-1.0597*	0.0094
ETHNIC	-0.4444*	0.0158
NONBRIT	-0.3276*	0.0166
ETHNBRIT	0.0101	0.0316
Family variables		
PCHILD	-0.1205*	0.0103
COHABIT	0.2329*	0.0116
MARRIED	0.0597*	0.0082
Property income variables		
PENSION	-0.5040*	0.0237
NONLABY	-0.00002*	0.0000
Years of education variables		
YED	0.2175*	0.0083
YED2	-0.0074*	0.0003
Educational attainment variables		
PGDEG	0.5698*	0.0241
DEG	0.3297*	0.0141
ALEVEL	-0.1016*	0.0123
NOQUAL	-0.5680*	0.0080
Regional variables		
SEAST	0.0061	0.0067
Log likelihood function	-115,454.20	
Restricted log likelihood	-165,334.20	
Model test	χ^2 = 99,760.03; df = 38; sig. = 0.0000	
N	247,774	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
* Significant at the 5% level

Table 5.4. Results of Model 2: probit estimates of participation equation [5.2] based on all individuals

In terms of the education variables, the relationship between years of education (YED, YED2) and an individual's propensity to participate in the labour market is n-shaped. Holding all other variables constant, individuals whose highest educational qualification is a first degree or a postgraduate degree are more likely to participate in the labour market than individuals with lower educational attainment. The opposite is true for individuals whose highest educational qualification is A level or for individuals with no qualifications at all. The negative signs on the co-efficients on the time trend variables (not shown – see Appendix 5.3) indicates that individuals in these quarters were less likely to participate in the labour market than other individuals, holding all other variables constant.

Turning to the results of the participation selection bias corrected estimates of the wage equation based on all workers with a nurse dummy variable; these are presented in Table 5.5. Identification of the wage equation is achieved through the omission of the non-labour income variable (NONLABY) which is statistically significant in the participation equation but not in the wage equation if it is re-estimated with no exclusion restrictions (see Appendix 5.6 Table A5.6.1). The proportion of variation in ln wages explained by the model is 30%, which is consistent with previous UK earnings function studies. The results are corrected for potential heteroscedasticity using Heckman's adjustment (Heckman, 1979). [19]

Both the co-efficients (the β's) and the marginal effects are reported. As discussed in Chapter 4 in this model the marginal effects on wages of variables that appear in both the participation selection equation and the wage equation consist of two components: the direct effect on the mean of lnW, which is the co-efficient in the wage equation β; and, the indirect effect on the mean of lnW exerted through the variable's influence on the selection bias correction term λ. Where a variable appears only in the wage equation but not in the participation equation the co-efficient in the wage equation β is also the marginal effect (there is no indirect effect).

The direct effect may be interpreted as an unbiased estimate of the impact of a variable on

[19] See Chapter 4 section 4.10. The correction is also applied to Models 4 and 5.

mean lnW that an individual in the population as a whole (including both workers and non-workers) can earn, on average (Killingsworth, 1983). The marginal effect also includes the indirect effect that quantifies the effect that a regressor has on the decision to participate and applies only to individuals who work.

On examination of the co-efficients in the wage equation (the direct effects) we can see that the co-efficient on NURSE (0.2007) shows that individuals employed as nurses earn on average 22% higher wages than individuals employed in other occupations, controlling for other measurable individual productive characteristics, and correcting for the decision to participate. This is very similar to the uncorrected estimates presented in Model 1.

The co-efficients on the years of education variables indicate that the relationship between years of education and earnings is non-linear (n-shaped) with maximum earnings occurring after 22 years of education. The co-efficients on the work experience variables are significant, as predicted in the Mincerian model, and also indicate an n-shaped experience-earnings profile. In this case earnings are maximised at 30 years of work experience.

In terms of educational attainment across all workers the transformed co-efficients suggest that the wage premium to obtaining a postgraduate degree is 33% and that for a first degree it is 26%. Having A levels as the highest educational qualification leads to a small positive wage premium (4%), and workers with no qualifications earn on average 5% lower wages than their better-educated counterparts. The positive co-efficient on DISABLE indicates that individuals in the whole population with health problems that affect paid work will in fact earn on average 16% higher wages than non-disabled workers, though the overall marginal effect (which includes the impact of disability on participation) is negative. The non-white

British and non-British whites in the population also receive positive wage premia of 7% and 11%, respectively, though the non-white non-British receive on average 8% lower earnings. Individuals living in the South East of England would earn on average 17% higher wages than those living outside of this area. Hourly wages increase with the number of hours worked (by an average of 0.09% per hour worked). Supervisors, managers or foremen earn on average 14% higher wages than their more junior counterparts, and the positive co-efficient on NWORKERS indicates that the larger the workplace the larger are average earnings. Workers on temporary contracts receive slightly higher earnings (by 2%) than those in permanent jobs.

Comparing the participation selection bias corrected wage equation co-efficients in Model 2 with the uncorrected co-efficients of Model 1 we can see there are some similarities (for example, the co-efficient on NURSE, and those on the years of experience variables, ETHNBRIT, SEAST, and the job characteristic variables). However, there are some notable differences also, especially pertaining to the years of education variables, the educational attainment variables and some of the personal characteristic variables. These are of a different order of magnitude to the uncorrected estimates of Model 1 and even in some cases (DISABLE and ETHNIC) of the opposite sign. The variables for which there is most discrepancy are those where an indirect effect is exerted through the participation equation.

In terms of the selection bias variable, λ, the co-efficient on λ is statistically significant and negative. As explained in Chapter 4 the negative co-efficient may be interpreted to mean that individuals who participate will earn a lower expected wage than (the same) individuals who do not participate would earn if they chose to participate. In other words the offered wage to non-participating individuals if they participated would be higher than that for participating individuals. This means that the reservation wage for non-participating individuals must be greater than that for participating individuals.

Turning now to the marginal effects presented in Table 5.5, it is interesting to note that for some variables that appear in both the participation equation and the wage equation (DISABLE and ETHNIC) the sign on the co-efficient (the direct effect) is the opposite to the sign of the total (direct plus indirect) marginal effect. This indicates that while the disabled might earn a higher wage than the non-disabled across the whole population, because being disabled has a strong negative on participation the marginal effect on wages is in fact negative. A similar effect is exerted by the ETHNIC variable. This interpretation is consistent with the results of the participation equation presented in Table 5.4 where the co-efficients on DISABLE and ETHNIC are statistically significant, negative and relatively large.

For other variables (YED, YED2, PGDEG, DEG, ALEVEL, NOQUAL and NONBRIT) the signs on the co-efficients and the marginal effects are the same but the magnitudes of the effects are somewhat different. For example, in the case of PGDEG the direct effect is less than the total marginal effect. This implies that there is a substantial wage premium to obtaining a postgraduate degree. It also means that obtaining a postgraduate degree exerts a positive influence on earnings indirectly through the decision to participate at all. The same may be said for the years of experience variables and for obtaining a first degree. In the case of having A levels as the highest education qualification we can see from Table 5.4 that this has a negative impact on participation. Therefore, while obtaining A levels would increase earnings on average in the whole population, for workers only this effect is muted by the negative impact on receiving a wage at all. A similar argument may also be made for NONBRIT.

In the case of ETHNBRIT and SEAST there is little difference between the co-efficient in the wage equation and the marginal effect. This indicates that the indirect effect of these variables is small, which is borne out by inspection of the magnitude of the relevant co-efficients in the participation equation (see Table 5.4).

	β^{1}	Std. Err.	Marginal Effects	Std. Err.
Constant	0.1898*	0.0320	0.1898*	0.0320
NURSE	0.2007*	0.0057	0.2007*	0.0057
Years of education variables				
YED	0.0946*	0.0042	0.1503*	0.0067
YED2	-0.0021*	0.0001	-0.0040*	0.0002
Educational attainment variables				
PGDEG	0.2851*	0.0084	0.4310*	0.0175
DEG	0.2283*	0.0054	0.3127*	0.0105
ALEVEL	0.0423*	0.0050	0.0162#	0.0093
NOQUAL	-0.0507*	0.0045	-0.1962*	0.0068
Work experience variables				
EXP	0.0297*	0.0005	0.0297*	0.0005
EXP2	-0.0005*	0.0000	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	0.1510*	0.0082	-0.1203*	0.0101
ETHNIC	0.0690*	0.0077	-0.0448*	0.0127
NONBRIT	0.1053*	0.0075	0.0214#	0.0130
ETHNBRIT	-0.0881*	0.0155	-0.0855*	0.0254
Region variables				
SEAST	0.1549*	0.0027	0.1565*	0.0050
Job characteristic variables				
HOURSPW	0.0009*	0.0001	0.0009*	0.0001
MANAGE	0.1350*	0.0027	0.1350*	0.0027
NWORKERS	0.1355*	0.0024	0.1355*	0.0024
TEMP	0.0153*	0.0043	0.0153*	0.0043
Selection bias variables				
λ	-0.4022*	0.0110		
Adjusted R^2			0.3049	
Model test			F(50, 151,893) = 1,334.19; p = 0.0000	
N			151,944	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table 5.5. Results of Model 2: participation selection bias corrected estimates of wage equation [5.3] based on all workers with NURSE dummy

Those variables that appear in both the participation equation and the wage equation in Model

2 have marginal effects that more closely resemble the co-efficient in the wage equation in Model 1 (as noted above the co-efficients on these variables in Model 2 were somewhat different to those in Model 1). Intuitively (and very crudely) this is because the co-efficients in Model 1 and the marginal effects in Model 2 both pertain to workers only (while the co-efficients in Model 2 pertain to the whole sample of workers and non-workers). It also indicates that the participation selection bias effect, while significant, is not great.

5.7.3. Model 3

In Models 1 and 2 we estimate a wage equation based on all workers in the sample and include a dummy variable for whether or not the worker is employed as a nurse. While this method allows estimation of the average wage premium to being employed as a nurse it makes the strong assumption that the impact on mean lnW of each of the other variables in the wage equation (such as whether or not the worker has a degree) are the same, regardless of whether the worker is employed as a nurse or in some other occupation. We now relax this restriction and allow for differences in co-efficients (the β's) between nurses and other workers in the sample. Model 3 estimates separate OLS wage equations for each occupation group (nurses and all other workers). The NURSE dummy variable is now omitted, but a dummy variable is included for whether or not the worker (nurse or otherwise) has a nursing qualification (NURSEQUA). This model makes no adjustment for selection bias. As with Model 1 this model is estimated using the sub-sample of individuals in the data who participate. The results of Model 3 are presented in Table 5.6.

The co-efficients in Model 3 are consistent with the Mincerian model of earnings, though there are differences between the co-efficients for nurses and all other workers. In terms of the years of education variables the earnings-years of education profile for both nurses and all other workers is n-shaped. For nurses the maximum earnings occur after 16 years of education. For all other workers the maximum occurs at 19 years. As before, the co-efficients on the work experience variables also indicate a concavity in the experience-earnings profile. In this case the relationship between work experience and earnings is also n-shaped with earnings maximised at 33 years of work experience for both nurses and all other workers.

In terms of the educational attainment variables the transformed co-efficients on NURSEQUA suggests that the wage premium to obtaining a nursing qualification is greater for nurses than other workers (27% versus 18%). Also as one might expect the benefits in terms of increased wages to having a postgraduate degree or first degree are, while positive, lower for nurses than other workers. Nurses with a postgraduate degree earn on average 25% higher wages than nurses without such a degree, whereas for workers in other occupations the premium is 48%. To a lesser extent the same is true for first degrees with the wage premium to nurses of having a degree at 10% and for other workers it is 35%. Nurses with A levels as the highest education attainment have on average 15% lower earnings than other nurses, and for all other workers the premium is +3%. The co-efficient on ALEVEL for nurses reflects the fact that nurses with A levels as their highest qualification are unlikely to have a nursing qualification. Unsurprisingly having no qualifications has a negative impact on earnings in both sectors. The impact is greater for individuals employed as nurses who receive 23% lower wages compared with their better-educated nurses, compared with an effect of –13% for all other workers.

In terms of the personal characteristic variables the co-efficient on DISABLE in the wage equation for nurses indicates that nurses with health problems that affect paid work will on average earn 4% lower wages than non-disabled nurses. For all other workers the effect is more pronounced, with disabled workers earning on average 8% less than the non-disabled. White non-British nurses earn higher (7%) wages than white British nurses, with a similar premium for white non-British workers in other occupations (5%).

Regional effects are less pronounced for nurses than all other workers. Nurses residing in the South East of England earn on average 7% higher wages than other nurses, while for all other workers the average effect is 18%.

In terms of the job characteristic variables, there are a number of differences between nurses and all other workers. First, nurses' hourly wages are inversely related to the number of hours worked (on average hourly wages decrease by 0.43% for every hour worked), but there is an opposite effect, on average, for all other workers (hourly wages increase by 0.09% for every hour worked). There is a wage premium to being employed in a managerial position for both nurses and all other workers (7% for nurses, 15% for other workers). For nurses, being employed in a relatively large workplace or being employed on a temporary contract are negatively related to earnings, while for all other workers on average these effects are positive.

An interesting result is apparent from the co-efficients on the time trend variables (not shown – see Appendix 5.3 Table A5.3.3). For all other workers the co-efficients on the time trend variables are as expected: they are generally statistically significant; have the expected sign (positive); and, are generally of the expected rank order of magnitude (i.e. generally the co-efficients increase with time, albeit in a non-uniform manner). For nurses the picture is somewhat different, the main point being that only in the later quarters are the co-efficients statistically significant. What this seems to indicate is that wage increases over time for nurses between quarters 0 and 26 (Winter 1992/3 to Summer 1999) were not statistically significant after controlling for measurable individual productive characteristics.

	Nurses		All Other Workers	
	β [1]	Std. Err.[2]	β [1]	Std. Err.[2]
CONSTANT	0.8263*	0.1498	-0.1135*	0.0462
Years of education variables				
YED	0.0975*	0.0196	0.1329*	0.0062
YED2	-0.0030*	0.0007	-0.0035*	0.0002
Educational attainment variables				
NURSEQUA	0.2389*	0.0214	0.1673*	0.0082
PGDEG	0.2252*	0.0422	0.3924*	0.0079
DEG	0.0998*	0.0164	0.2995*	0.0053
ALEVEL	-0.1683*	0.0597	0.0260*	0.0052
NOQUAL	-0.2666*	0.0904	-0.1355*	0.0033
Work experience variables				
EXP	0.0197*	0.0018	0.0327*	0.0005
EXP2	-0.0003*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	-0.0435#	0.0224	-0.0830*	0.0055
ETHNIC	-0.0262	0.0240	-0.0118	0.0076
NONBRIT	0.0659*	0.0201	0.0523*	0.0083
ETHNBRIT	0.0005	0.0423	-0.1063*	0.0172
Regional variables				
SEAST	0.0707*	0.0106	0.1613*	0.0026
Job characteristic variables				
HOURSPW	-0.0043*	0.0005	0.0009*	0.0001
MANAGE	0.0670*	0.0122	0.1365*	0.0029
NWORKERS	-0.0402*	0.0119	0.1386*	0.0025
TEMP	-0.0790*	0.0262	0.0150*	0.0053
Adjusted R^2	0.1485		0.2936	
Model test	$F(49, 6{,}558) = 24.51$; $p = 0.0000$		$F(49, 145{,}285) = 1{,}233.62$; $p = 0.0000$	
N	6,608		145,335	

[1] Dependent variable is LNWAGE
[2] Results corrected for heteroscedasticity using White's estimator
* Significant at the 5% level
\# Significant at the 10% level

Table 5.6. Results of Model 3: OLS estimates of separate wage equations [5.4] and [5.5] for nurses and all other workers

5.7.4. Model 4

In Model 4 we estimate wage equations separately for nurses and all other workers with corrections for participation selection bias. Turning first to the participation equations, we estimate separate equations for nurses and all other workers. In this model it is necessary to obtain information on non-participating nurses (and also non-participating non-nurses). As explained above the QLFS dataset does not collect information on occupation from respondents if they are not participating in the labour market. We therefore define non-participating nurses as individuals who have a nursing qualification but who are not participating in the labour market. Non-participating non-nurses are then defined as all other non-participators. An alternative specification of the participation equation is possible if we are prepared to accept that non-participating nurses are the same as non-participating non-nurses (i.e. that the participation equations for the two groups are the same). In this case the participation equation may be estimated using the whole sample of individuals in the data (all participators and non-participators). This is identical to the participation equation used in Model 2 and presented in Table 5.4. $\lambda(p)$ may be estimated for each observation in the whole sample. The separate wage equations for nurses and all other workers may then be estimated using the value of $\lambda(p)$ for each relevant sub-sample of workers.

It is possible to test the hypothesis that nurses and all other workers as defined have different participation equations. This is equivalent to a Chow-type test (which we use below in the context of the wage equations) applied to a probit model based on the likelihood ratio (LR) statistic (see Greene, 2000, for a discussion of restricted log-likelihoods and chow-type tests for probit models). The statistic is:

$$LR = -2(\ln \hat{L}_r - \ln \hat{L}_u) \qquad [5.23]$$

where $\hat{L}_r$ and $\hat{L}_u$ are the log-likelihood functions evaluated at the restricted and unrestricted estimates, respectively.

The null hypothesis is that the co-efficients of the probit model of participation for nurses and all other workers are the same. The alternative hypothesis is that an altogether different participation equation applies for the two groups of individuals (nurses and all other workers). To test for this we use the probit counterpart to the Chow test. The restricted model in this instance is based on all 247,774 observations in the data. The log-likelihood for the participation equation in this model is –115,454.20. The log-likelihoods for this model based on the 8,878 observations for nurses only and the 238,896 observations for all other workers only are –3,470.225 and –111,627.8, respectively. Therefore the log-likelihood for the unrestricted model with separate equations is the sum, -115,098. 025. The χ^2 squared statistic for testing the 38 restrictions of the pooled model is twice the difference between the restricted and unrestricted log-likelihoods (see equation [5.23]), or 716.35. The 95% critical value from the χ^2 squared distribution is approximately 50.00. So, at this significance level the null hypothesis that the constant terms and the co-efficients are the same on the probit model of participation for nurses and all other workers is rejected. The conclusion is that it is appropriate to estimate participation equations for nurses and all other workers separately using the methods described.

The results of the separate participation equations estimated for nurses and all other workers are presented in Table 5.7.[20]

	Nurses		All Other Workers	
	δ [1]	Std.Err.	δ [1]	Std.Err.
Constant	1.6757*	0.5629	-1.1004*	0.0729
Age variables				
AGE	0.0536*	0.0156	0.0715*	0.0020
AGE2	-0.0009*	0.0002	-0.0008*	0.0000
Personal characteristic variables				
DISABLE	-1.2072*	0.0514	-1.0545*	0.0096
ETHNIC	0.0763	0.0950	-0.4564*	0.0161
NONBRIT	0.1245	0.0878	-0.3457*	0.0170
ETHNBRIT	-0.1727	0.1788	0.0117	0.0323
Family variables				
PCHILD	-0.0941	0.0575	-0.1297*	0.0105
COHABIT	0.3699*	0.0900	0.2316*	0.0117
MARRIED	-0.2699*	0.0466	0.0718*	0.0084
Property income variables				
PENSION	-0.8275*	0.1024	-0.4846*	0.0244
NONLABY	-0.0001*	0.0000	-0.00007*	0.0000
Years of education variables				
YED	0.0343	0.0587	0.2135*	0.0084
YED2	-0.0025	0.0019	-0.0073*	0.0003
Educational attainment				
PGDEG	0.2900#	0.1574	0.5868*	0.0245
DEG	0.0897	0.0751	0.3443*	0.0145
ALEVEL	[2]	[2]	-0.0932*	0.0123
NOQUAL	[2]	[2]	-0.5638*	0.0081
Regional variables				
SEAST	-0.1677*	0.0390	0.0125#	0.0068
Log likelihood function	-3,470.225		-111627.8	
Restricted log likelihood	-5,047.129		-159934.6	
Model test	χ^2 = 3,153.749; df = 36; sig. = 0.0000		χ^2 = 96613.52; df = 38; sig. = 0.0000	
N	8,878		238,896	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[2] ALEVEL and NOQUAL predict PART perfectly for nurses and so are omitted
* Significant at the 5% level
# Significant at the 10% level

Table 5.7. Results of Model 4: probit estimates of participation equations [5.6] and [5.8] estimated separately for nurses and all other workers

In terms of the model of individual labour supply the key variables in the participation equation are the property income variables. For both nurses and all other workers the co-

[20] For comparative purposes the alternative to Model 4 estimated assuming the same participation equation for nurses and all other workers (called Model 4a) is presented in Appendix 5.7.

efficients on these variables are statistically significant and of the expected sign (negative), reflecting the fact that as property income increases the propensity to participate in the labour market decreases, all other things being equal.

There are differences, in terms of the co-efficients on the other variables, between nurses and other workers. Most notably for all other workers ETHNIC, NONBRIT and PCHILD are statistically significant and negatively related to probability of participation, while for nurses they are not significant. Years of education is not significantly related to participation in the labour market for nurses, though it is for other workers, and living in the South East has a negative impact on participation for nurses and a positive impact of other workers. For all other variables the co-efficients are of the same sign and significance (though not necessarily of the same order of magnitude). The results for all other workers in model 4 reflect more closely the results for all individuals in Model 2. This is unsurprising since working and non-working nurses form only 4% of the total population in the data.

Turning to the results of the participation selection bias corrected estimates of the wage equation estimated separately for nurses and all other workers; these are presented in Table 5.8. The wage equations are identified by the property income variables in the occupation selection equation (see Appendix 5.6 Table A5.6.2).

The direct effects (the co-efficients) and the marginal effects have a similar interpretation as before in Model 2. The direct effect may be interpreted as an unbiased estimate of the impact of a variable on mean lnW that an individual in that occupation group (nurses or other workers) as a whole (including both participators and non-participators) can earn, on average. The marginal effect also includes the indirect effect that quantifies the effect that a regressor has on the decision to participate and applies only to individuals within each occupation group who work.

The co-efficients on the years of education variables indicate that the relationship between years of education and earnings for all workers is non-linear (n-shaped). For nurses the maximum occurs after 16 years of education. For other workers the maximum occurs after 23 years. Only a very small proportion of non-nurses in the data actually obtained this number of years of education – less than 1% of the sample. This is expected given that only around 3% of non-nurses have a postgraduate qualification (see Table 5.1). The co-efficients on the work experience variables are also significant. In this case earnings are maximised at 33 years of work experience for nurses and 30 years for other workers.

In terms of educational attainment the transformed co-efficients suggest that the premium to obtaining a nursing qualification is higher for nurses than other workers (27% versus 17%). The wage premium to obtaining a postgraduate degree is 25% for nurses and 33% for other workers. The premium to obtaining a first degree is also lower for nurses (11% versus 27%). Having A levels as the highest educational qualification leads to a negative wage premium for nurses (-15%) and small positive wage premium (5%), for other workers. Nurses with no qualifications earn on average 23% lower wages than their better-educated counterparts, while for other workers with no qualifications the effect is –4%.

For nurses the co-efficients on being disabled, non-white or both non-white and non-British are not statistically significant, while being non-British has a positive impact on wages (7%), controlling for all other variables. For other worker these variables are significant.

Nurses living in the South East of England earn on average 7% higher wages than nurses living outside of this area. For other workers the premium is even higher at 17%. In terms of the job characteristic variables, there are a number of differences between nurses and all other workers. First, nurses' hourly wages are inversely related to the number of hours worked but there is an opposite effect, on average, for all other workers. There is a wage premium to being employed in a managerial position for both nurses and all other workers (7% for nurses, 14% for other workers). For nurses, being employed in a relatively large workplace or being employed on a temporary contract are negatively related to earnings, while for all other workers on average these effects are positive.

	Nurses				All Other Workers			
			Marginal				Marginal	
	β [1]	Std.Err.	Effects	Std.Err.	β [1]	Std.Err.	Effects	Std.Err.
Constant	0.8233*	0.1461	0.8233*	0.1461	0.1853*	0.0324	0.1853*	0.0324
Years of education variables								
YED	0.0979*	0.0190	0.0978*	0.0419	0.0931*	0.0042	0.1464*	0.0068
YED2	-0.0031*	0.0006	-0.0030*	0.0014	-0.0020*	0.0001	-0.0038*	0.0002
Educational attainment variables								
NURSEQUA	0.2388*	0.0186	0.2388*	0.0186	0.1608*	0.0075	0.1608*	0.0075
PGDEG	0.2257*	0.0382	0.2247*	0.1073	0.2870*	0.0086	0.4335*	0.0178
DEG	0.1000*	0.0186	0.0997*	0.0513	0.2362*	0.0056	0.3221*	0.0108
ALEVEL	-0.1681*	0.0469	-0.1681*	0.0469	0.0496*	0.0051	0.0263*	0.0093
NOQUAL	-0.2665*	0.0675	-0.2665*	0.0675	-0.0470*	0.0045	-0.1877*	0.0068
Work experience variables								
EXP	0.0197*	0.0019	0.0197*	0.0019	0.0298*	0.0005	0.0298*	0.0005
EXP2	-0.0003*	0.0001	-0.0003*	0.0001	-0.0005*	0.0000	-0.0005*	0.0000
Personal characteristic variables								
DISABLE	-0.0468	0.0298	-0.0426	0.0443	0.1419*	0.0083	-0.1213*	0.0103
ETHNIC	-0.0261	0.0244	-0.0264	0.0652	0.0738*	0.0079	-0.0401*	0.0129
NONBRIT	0.0660*	0.0220	0.0656	0.0601	0.1090*	0.0078	0.0227#	0.0133
ETHNBRIT	0.0004	0.0455	0.0010	0.1226	-0.0925*	0.0161	-0.0896*	0.0261
Regional variables								
SEAST	0.0704*	0.0106	0.0709*	0.0270	0.1572*	0.0027	0.1603*	0.0051
Job characteristic variables								
HOURSPW	-0.0043*	0.0004	-0.0043*	0.0004	0.0010*	0.0001	0.0010*	0.0001
MANAGE	0.0670*	0.0110	0.0670*	0.0110	0.1333*	0.0028	0.1333*	0.0028
NWORKERS	-0.0400*	0.0121	-0.0400*	0.0121	0.1386*	0.0024	0.1386*	0.0024
TEMP	-0.0791*	0.0190	-0.0791*	0.0190	0.0157*	0.0044	0.0157*	0.0044
Selection bias variables								
λ	0.0055	0.0361			-0.3920*	0.0112		
Adjusted R^2	0.1483				0.2997			
Model test	$F(50, 6{,}557) = 24.02$; p value = 0.0000				$F(50, 145285) = 1245.23$; p value = 0.0000			
N	6,608				145,336			

[1] Dependent variable is LNWAGE
* Significant at the 5% level
[#] Significant at the 10% level

Table 5.8. Results of Model 4: participation selection bias corrected estimates of wage equations [5.7] and [5.9] estimated separately for nurses and all other workers

A similar result is apparent from the time trend variables (not shown – see Appendix 5.3) as in Model 3: wage increases for nurses over time were not statistically significant after controlling for measurable individual productive characteristics until the late 1990s.

For nurses the co-efficient on the selection bias correction term is numerically very small and not statistically significant. This is interpreted to mean that participation selection bias is not significant for this group. This also explains why there is little difference between the co-efficients and the full marginal effects for nurses reported in Table 5.8 (the indirect effects exerted through λ are small and/or insignificant). For all other workers, however, the co-efficient on λ is negative and statistically significant. This means that individuals who participate in this group will earn lower expected wages than (the same) individuals who do not participate would earn if they chose to participate. For all other workers, both the interpretation of λ and the interpretation of the indirect effects and full marginal effects are the same as those in Model 2.

5.7.5. Model 5

Model 5 estimates an occupation selection bias corrected wage equation separately for nurses only and for all other workers only. The results of the occupation selection equation are presented in Table 5.9. These are the results of a probit model estimating the likelihood that a worker will be employed as a nurse based on the sample of all workers. In this probit model the dependent variable is dichotomous taking the value one if the worker is employed as a nurse and zero if the worker is employed in some other occupation. The key independent variables in the occupation selection equation relate to individual productive characteristics measured via education and training variables. NURSEQUA is a dummy variable measuring whether or not the individual has a nursing qualification. The co-efficient on NURSEQUA is statistically significant and positive confirming the view that having a nursing qualification is a basic requirement for being employed as a nurse. Other education and training variables that are included are years of education variables (YED and YED2), which are insignificant, and educational attainment variables (PGDEG, DEG, ALEVEL and NOQUAL).

	γ [1]	Std. Err.
Constant	-2.7510*	0.3250
NURSEQUA	3.0550*	0.0227
Age variables		
AGE	0.0012	0.0074
AGE2	-0.0001	0.0001
Personal characteristic variables		
DISABLE	-0.1554*	0.0470
ETHNIC	0.2124*	0.0537
NONBRIT	0.3157*	0.0501
ETHNBRIT	0.0562	0.1037
Family variables		
PCHILD	0.0937*	0.0321
COHABIT	-0.0666#	0.0376
MARRIED	-0.1150*	0.0274
Property income variables		
PENSION	-0.0404	0.0964
NONLABY	-0.00003*	0.0000
Years of education variables		
YED	0.0452	0.0388
YED2	-0.0022#	0.0013
Educational attainment variables		
PGDEG	-0.5472*	0.0738
DEG	-0.0804*	0.0398
ALEVEL	0.1444*	0.0480
NOQUAL	-0.3265*	0.0598
Region variables		
SEAST	-0.0980*	0.0227
Log likelihood function	-9,528.88	
Restricted log likelihood	27,179.79	
Model test	$\chi^2 = 35{,}301.81$; df = 50; sig. = 0.0000	
N	151,944	

[1] Dependent variable is whether the individual is employed as a nurse (NURSE = 1) or not (NURSE = 0).
* Significant at the 5% level
[#] Significant at the 10% level

Table 5.9. Results of Model 5: probit estimates of occupational selection equation [5.10] based on all workers

The co-efficients on PGDEG and DEG are negative which indicates that workers with degrees (postgraduate and undergraduate) are less likely to be employed as nurses. Individuals with A levels as their highest qualification are more likely to be employed as nurses, reflecting the fact that A levels are often seen as the minimum entrance requirement into nurse training college. Having no education qualifications reduces the likelihood of being employed as a nurse.

In terms of the other variables included in the occupation selection equation that may have an impact on the decision to become a nurse important variables are the existence of health problems affecting paid work, cohabiting or being married, the magnitude of non-labour income and whether or not the individual resides in the South East of England. These factors reduce the likelihood that a worker chooses to be employed as a nurse. Being non-white or non-British and having children increases the likelihood of a worker being employed as a nurse.

We turn now to the results of the separate occupation selection bias corrected wage equation estimates for nurses and all other workers, presented in Table 5.10. Both the co-efficients (the β's) and the marginal effects are reported. Identification of the wage equation is achieved through the omission of the non-labour income variable (NONLABY) – see Appendix 5.6 Table A5.6.3).

On examination of the co-efficients in the wage equation (the β's) we can see that just as with all the previous models the co-efficients in Model 5 are consistent with the Mincerian model of earnings, though there are differences between the co-efficients for nurses and all other workers. In terms of the years of education variables the earnings-years of education profile for both nurses and all other workers is n-shaped. For nurses the maximum earnings occur after 16 years of education, and for all other workers the maximum occurs at 19 years (both of these are the same as in Model 3). As before, the co-efficients on the work experience variables also indicate a concavity in the experience-earnings profile. In this case the relationship between work experience and earnings is also n-shaped with earnings maximised at 32 and 33 years of work experience for nurses and all other workers, respectively.

In terms of the educational attainment variables the transformed co-efficients on NURSEQUA suggests that the wage premium to obtaining a nursing qualification is greater for nurses than other workers (92% versus 25%). The benefits in terms of increased wages to having a postgraduate degree or first degree are, while positive, lower for nurses than other workers (returns to a postgraduate degree are 17% for nurses and 48% for workers in other occupations, and for a first degree they are 10% and 35%, respectively). Nurses with A levels as the highest education attainment have on average 13% lower earnings than other nurses, and for all other workers the premium is +3%. Having no qualifications has a negative impact on earnings in both sectors. The impact is greater for individuals employed as nurses who receive 28% lower wages compared with their better-educated nurses, compared with an effect of –13% for all other workers.

In terms of the personal characteristic variables the co-efficient on DISABLE in the wage equation for nurses indicates that nurses with health problems that affect paid work will on average earn 6% lower wages than non-disabled nurses. For all other workers the effect is more pronounced, with disabled workers earning on average 8% less than the non-disabled. White non-British nurses earn 10% higher wages than white British nurses, with a slightly smaller premium for white non-British workers in other occupations (5%).

Regional effects are less pronounced for nurses than all other workers. Nurses residing in the South East of England earn on average 6% higher wages than other nurses, while for all other workers the average effect on lnW is 17%.

In terms of the job characteristic variables, nurses' hourly wages are inversely related to the number of hours worked, but there is an opposite effect, on average, for all other workers. There is a wage premium to being employed in a managerial position for both nurses and all other workers (6% and 15%, respectively). For nurses, being employed in a relatively large workplace or being employed on a temporary contract are negatively related to earnings, while for all other workers on average these effects are positive.

As in Models 3 and 4 for all other workers the co-efficients on the time trend variables are generally statistically significant; have the expected sign (positive); and, are generally of the expected rank order of magnitude (i.e. generally the co-efficients increase with time, albeit in a non-uniform manner) (data not shown – see Appendix 5.3). For nurses again the picture is somewhat different; really only in the later quarters are the co-efficients statistically significant.

For nurses the co-efficient on the selection bias variable is statistically significant and positive, while for non-nurses it is not statistically significant. As explained above the interpretation of the selection bias effect is ambiguous.

We turn now to the marginal effects presented in Table 5.10. The total marginal effect of a variable in the wage equation that also appears in the occupation selection equation consists of two components: the direct effect on the mean of lnW, which is the co-efficient in the wage equation β; and, the indirect effect on lnW through the influence on the selection bias variable λ. Where a variable appears only in the wage equation but not in the occupation selection equation the co-efficient in the wage equation β is also the marginal effect (there is no indirect effect). The co-efficient is an unbiased estimate of the return to a variable that an individual in the self-selected sub-sample (for example, nurses when Nu = 1) can obtain, on average. The indirect effect captures the impact of the variable on the change in the probability of choosing to be employed as a nurse. As noted in Chapter 4 the indirect effect is subtracted from the direct effect to give the total effect, leaving only the marginal effect of a change given that Nu = 1 (say) to begin with. In the wage equation for nurses the sign of the indirect effect of a variable is the same as the sign on the co-efficient of that variable in occupation selection equation. This is because nurses on average earn higher incomes than other workers and so any variable that has a positive effect on the decision to become a nurse has a positive indirect effect and vice versa. As noted in Chapter 4 we subtract the indirect effect to measure the change in the conditional expectation of $\ln W_n$ given the decision to be employed as a nurse with respect to a change in the variable of interest. For example, the marginal effect of obtaining a nursing qualification is less than the co-efficient. From Table 5.9 we know that having a nursing qualification is positively related to being employed as a nurse (i.e. the indirect effect is also positive). Subtracting this positive indirect effect from the co-efficient (as delineated in equation [4.58]) results in a smaller marginal effect. Following this line of reasoning the indirect effects are of the expected sign and the marginal effects are related in the predicted way to the co-efficients in all instances.

For all the variables considered (save some of the time trend variables) the sign on the co-efficient is the same as the sign on the marginal effect. However, for some variables (most noticeably NURSEQUA) the sign on the co-efficient and the marginal effect are the same but the magnitudes of the effects are somewhat different. This indicates that NURSEQUA exerts a strong indirect effect (positive for nurses, negative for all other workers). This interpretation is consistent with the results of the occupation selection equation presented in Table 5.9 where for the decision to be employed as a nurse the co-efficient on NURSEQUA is statistically significant, positive and large.

In the case of many of the other variables that appear in both the occupation selection equation and the wage equation (for example, DEG and ALEVEL) there is little difference between the co-efficient in the wage equation and the marginal effect. This indicates that the indirect effect of these variables is small, which is borne out by inspection of the magnitude of the relevant co-efficients in the occupation selection equation (see Table 5.9).

	β [1]	Std. Err.	Nurses Marginal Effects	Std. Err.	β [1]	Std. Err.	All Other Workers Marginal Effects	Std. Err.
ONE	0.2203	0.3002	0.2203	0.3002	-0.1142*	0.0299	-0.1142*	0.0299
Years of education variables								
YED	0.1100*	0.0199	0.1050*	0.0317	0.1332*	0.0039	0.1315*	0.0250
YED2	-0.0035*	0.0007	-0.0032*	0.0010	-0.0035*	0.0001	-0.0034*	0.0008
Educational attainment variables								
NURSEQUA	0.6542*	0.1803	0.3144#	0.1803	0.2240*	0.0520	0.1126*	0.0520
PGDEG	0.1587*	0.0481	0.2196*	0.0673	0.3911*	0.0073	0.4110*	0.0476
DEG	0.0920*	0.0192	0.1010*	0.0318	0.2992*	0.0048	0.3021*	0.0258
ALEVEL	-0.1415*	0.0470	-0.1576*	0.0560	0.0263*	0.0046	0.0210	0.0309
NOQUAL	-0.3274*	0.0703	-0.2910*	0.0799	-0.1360*	0.0035	-0.1241*	0.0382
Work experience variables								
EXP	0.0194*	0.0019	0.0194*	0.0019	0.0327*	0.0005	0.0327*	0.0005
EXP2	-0.0003*	0.0001	-0.0003*	0.0001	-0.0005*	0.0000	-0.0005*	0.0000
Personal characteristic variables								
DISABLE	-0.0612*	0.0224	-0.0439	0.0374	-0.0834*	0.0051	-0.0777*	0.0303
ETHNIC	-0.0073	0.0263	-0.0309	0.0431	-0.0112	0.0072	-0.0189	0.0349
NONBRIT	0.0967*	0.0263	0.0616	0.0413	0.0533*	0.0072	0.0418	0.0327
ETHNBRIT	0.0079	0.0468	0.0016	0.0809	-0.1059*	0.0156	-0.1080	0.0678
Regional variables								
SEAST	0.0592*	0.0117	0.0701*	0.0186	0.1611*	0.0025	0.1646*	0.0147
Job characteristic variables								
HOURSPW	-0.0042*	0.0004	-0.0042*	0.0004	0.0009*	0.0001	0.0009*	0.0001
MANAGE	0.0672*	0.0110	0.0672*	0.0110	0.1365*	0.0028	0.1365*	0.0028
NWORKERS	-0.0363*	0.0121	-0.0363*	0.0121	0.1386*	0.0024	0.1386*	0.0024
TEMP	-0.0785*	0.0190	-0.0785*	0.0190	0.0149*	0.0044	0.0149*	0.0044
Selection bias variables								
λ	0.1747*	0.0755			-0.0573	0.0520		
Adjusted R^2	0.1490				0.2936			
Model test	$F(50, 6{,}557) = 24.14$; $p = 0.0000$				$F(50, 145{,}284) = 1{,}208.98$; $p = 0.0000$			
N	6,608				145,335			

[1] Dependent variable is LNWAGE
* Significant at the 5% level
\# Significant at the 10% level

Table 5.10. Results of Model 5: occupation selection bias corrected wage equations [5.11] and [5.13] estimated separately for nurses and all other workers

5.8. Comparing wage equations for nurses and all other workers: the Chow test

So far in this chapter we have constructed five statistical models to analyse the determinants of nurses' earning. These analyses have been based on extended earnings functions for nurses and other workers, correcting for potential participation selection bias and occupation selection bias. In numerical terms there are differences between the co-efficients of the wage equations for nurses and other workers estimated by Models 3-5. We now wish to analyse in a statistical way whether or not these differences are significant.

The following is based on Gujarati (1988) and Greene (2000). Formally, we wish to ascertain whether the relationship between lnW and X is different for nurses and all other workers. We would also like to know whether if there is a difference it is in the intercepts, the slopes, or both. Suppose we estimate the following wage equations:

For nurses: $$\ln W_i = \alpha_n + \beta_n X_{ni} + U_{ni} \qquad [5.24]$$

For all other workers: $$\ln W_i = \alpha_o + \beta_o X_{oi} + U_{oi} \qquad [5.25]$$

For simplicity the selection bias correction terms are omitted from the exposition (but not from the analysis). Comparing these regressions may yield one of four outcomes:

1. $\alpha_n = \alpha_o$ and $\beta_n = \beta_o$: the two regressions are identical (the wage equations for nurses and all other workers are the same);
2. $\alpha_n \neq \alpha_o$ and $\beta_n = \beta_o$: the two regressions differ only in their intercepts (they are parallel regressions);
3. $\alpha_n = \alpha_o$ and $\beta_n \neq \beta_o$: the two regressions have the same intercepts but have different slopes (they are concurrent regressions); or,
4. $\alpha_n \neq \alpha_o$ and $\beta_n \neq \beta_o$: the two regressions have different intercepts and slopes (they are dissimilar regressions).

It is possible to test for these differences across two regressions using the Chow test (Chow, 1960). The general principle underlying the Chow test is to apply an F test as a test of structural change. In specifying a regression model we assume that it's co-efficients apply to all the observations in the sample (e.g. all workers). In the Chow test we test the hypothesis that some or all of the regression co-efficients are different in different subsets of the data (e.g. for nurses and other workers). We first construct an unrestricted regression model that allows the intercept terms, some or all of the slope co-efficients, or both of these to be different in different subsets of the data. We then compare this using an F test to a restricted model that has linear restrictions imposing homogenous intercept terms and slope co-efficients. The Chow test is based on the following assumptions:

1. $U_{ni} \sim N(0,\sigma^2)$, $U_{oi} \sim N(0,\sigma^2)$; and,
2. U_{ni} and U_{oi} are distributed independently.

We first use the most general form of the Chow test to ascertain whether there are any structural differences between the two regressions (that is, to see whether the wage equations for nurses and all other workers are the same). The test is conducted as follows:

1. Combine observations for nurses and all other workers and estimate the following single pooled regression based on all workers:

Pooled regression (all workers): $\ln W_i = \alpha + \beta X_i + U_i$ [5.26]

This is the restricted regression model. From this regression obtain the residual sum of squares (RSS) with degrees of freedom $N_1 + N_2 - k$, where N_1 = number of nurses in the sample, N_2 = number of workers in occupations other than nursing in the sample and k = number of parameters. This is the restricted residual sum of squares, RSS_R.

2. Run the two individual regressions delineated in equations [5.24] and [5.25] and obtain their RSS (call these RSS_n and RSS_o for nurses and all other workers, respectively) with degrees of freedom $N_1 - k$ and $N_2 - k$, respectively. This is the unrestricted regression model. Add $RSS_n + RSS_o$ to obtain the unrestricted residual sum of squares, RSS_U.

3.Calculate the following test statistic for testing the restrictions that the two regressions for nurses and all other workers are the same:

$$F_{J, N_1 + N_2 - 2k} = \frac{(RSS_R - RSS_U)/J}{RSS_U/(N_1 + N_2 - 2k)} \qquad [5.27]$$

where J = number of linear restrictions. If this computed F exceeds the critical F, reject the null hypothesis that the regression models are the same across the two subsets of the data.

We conduct this general Chow test for structural change using wage equations uncorrected for selection bias and also using participation selection bias and occupation selection bias corrected estimates.

5.8.1. No correction for selection bias

This test is based on the results of Model 3. OLS earnings functions are estimated separately for nurses and all other workers (equations [5.24] and [5.25]), using the same regressors as

Model 3 (see Table 5.6). The pooled regression is estimated on all workers using the same regressors. There is no correction for selection bias.

Test statistic:

$$F_{50,\,151{,}843} = \frac{(28007.0760 - 277866.4678)/50}{277866.4678/151{,}843} = 15.3234$$

Critical $F_{50,\,151{,}843}{}^{0.05} \approx 1.46$

15.3234 > 1.46, therefore reject the hypothesis that the uncorrected wage equations for nurses and all other workers are the same.

5.8.2. Participation selection bias corrected estimates

This test is based on the results of Models 2 and 4. The results of Model 4 are used as the unrestricted regression. The pooled (restricted) regression is estimated as in Model 2 but without the NURSE dummy).

Test statistic:

$$F_{51,\,151{,}841} = \frac{(27861.1249 - 25606.5769)/51}{25606.5769/151{,}841} = 262.1361$$

Critical $F_{50,\,151{,}843}{}^{0.05} \approx 1.45$

262.1361 > 1.45, therefore reject the null hypothesis that the participation selection bias corrected wage equations for nurses and all other workers are the same.

5.8.3. Occupation selection bias corrected estimates

This test is based on the results of Model 5. The results of Model 5 are used as the unrestricted regression. The pooled (restricted) regression is estimated on all workers using the same regressors.

Test statistic:

$$F_{51,\ 151{,}841} = \frac{(28004.5279 - 27859.2292)/51}{27859.2292/151{,}841} = 15.5279$$

Critical $F_{50,\ 151{,}843}^{\ 0.05} \approx 1.45$

15.5279 > 1.45, therefore reject the null hypothesis that the participation selection bias corrected wage equations for nurses and all other workers are the same.

In all three instances we reject the hypothesis that the wage equations for nurses and all other workers are the same. The wage equations for nurses and all other workers are not the same – the restricted regression model is too serious to impose.

5.8.4. Chow tests on sets of variables

The general formulation of the Chow test above lends itself to a number of variations that allow a wide range of related tests. We now focus on specific variables or sets of variables in the regression models to ascertain whether the structural differences between the wage equations of nurses and all other workers shown above arise from a difference in the intercept terms, differences in the slope co-efficients, or both.

We adopt a dummy variable approach which is slightly different to the one used above by first introducing a NURSE dummy variable to delineate differences in intercepts or slope co-efficients between the two wage equations. Suppose we combine observations for nurses and all other workers and estimate the following single pooled regression based on all workers:

$$\ln W_i = \alpha_1 + \alpha_2 Nu_i + \beta_1 X_i + \beta_2 (Nu_i X_i) + U_I \qquad [5.28]$$

where Nu is a dummy variable taking the value 1 if the individual is employed as a nurse and 0 if the individual is employed in some other occupation. Depending on the value of Nu we obtain the following:

$$E(\ln W_i \mid Nu = 1, X_i) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_i \qquad [5.29]$$

$$E(\ln W_i \mid Nu = 0, X_i) = \alpha_1 + \beta_1 X_i \qquad [5.30]$$

These are the separate wage equations for nurses and all other workers and are the same as equations [5.24] and [5.25], respectively, with $\alpha_n = (\alpha_1 + \alpha_2)$, $\alpha_o = \alpha_1$, $\beta_n = (\beta_1 + \beta_2)$ and $\beta_o =$

β_1. As before we compare using an F test an unrestricted and restricted regression model. The unrestricted model assumes that Nu = 1 for the intercept term and then for various slope co-efficients (effectively allowing differences in the wage equations for nurses and all other workers). The restricted model assumes that Nu = 0, which implies that the intercept term or slope co-efficient is the same for nurses and all other workers (and therefore that the wage equations are the same). By conducting this test across a range of variables we can ascertain where the structural differences supported by the general Chow tests described above originate. The unrestricted and restricted models that are compared using the F test are presented in Table 5.11.

Restrictions pertaining to	Unrestricted model [1,2,3]
NURSE dummy (intercept term)	$X_i + Nu_i$
Years of education variables	$X_i + Nu_i.YED_i + Nu_i.YED2_i$
Educational attainment variables	$X_i + Nu_i.PGDEG_i + Nu_i.DEG_i + Nu_i.ALEVEL_i + Nu_i.NOQUAL_i$
Work experience variables	$X_i + Nu_i.EXP_i + Nu_i.EXP2_i$
Personal characteristic variables	$X_i + Nu_i.DISABLE_i + Nu_i.ETHNIC_i + Nu_i.NONBRIT_i + Nu_i.ETHNBRIT_i$
Regional variables	$X_i + Nu_i.SEAST_i$
Job characteristic variables	$X_i + Nu_i.HOURSPW_i + Nu_i.MANAGE_i + Nu_i.NWORKERS_i + Nu_i.TEMP_i$

[1] The dependent variable is $\ln W_i$

[2] X_i = f(constant, YED_i, $YED2_i$, $PGDEG_i$, DEG_i, $ALEVEL_i$, $NOQUAL_i$, EXP_i, $EXP2_i$, $DISABLE_i$, $ETHNIC_i$, $NONBRIT_i$, $ETHNBRIT_i$, $SEAST_i$, $HOURSPW_i$, $MANAGE_i$, $NWORKERS_i$, $TEMP_i$, $Q1_i$, $Q2_i$, $Q3_i$, $Q4_i$, Q5, $Q6_i$, $Q7_i$, $Q8_i$, $Q9_i$, $Q10_i$, $Q11_i$, $Q12_i$, $Q13_i$, $Q14_i$, $Q15_i$, $Q16_i$, $Q17_i$, $Q18_i$, $Q19_i$, $Q20_i$, $Q21_i$, $Q22_i$, $Q23_i$, $Q24_i$, $Q25_i$, $Q26_i$, $Q27_i$, $Q28_i$, $Q29_i$, $Q30_i$, $Q31_i$). X_i also includes a selection bias correction term λ where relevant

[3] In all cases the restricted model assumes that $Nu_i = 0$

Table 5.11. Chow tests for differences in wage equations between nurses and all other workers

The test statistic for testing the restrictions that the two regressions for nurses and all other workers are the same is calculated using equation [5.27]. Results are estimated using wage equations uncorrected for selection bias and also using participation selection bias and occupation selection bias corrected estimates (see Table 5.12).

Restrictions pertaining to	RSS_U	RSS_R	Test statistic
No correction for selection bias (Model 3)			
NURSE dummy (intercept term)	28128.0694	28368.2288	$F_{1, 151,894} = 1296.8813$*
Years of education variables	28105.4621	28368.2288	$F_{2, 151,893} = 710.0473$*
Educational attainment variables	28364.8911	28368.2288	$F_{4, 151,891} = 4.4683$*
Work experience variables	28180.3400	28368.2288	$F_{2, 151,893} = 506.3636$*
Personal characteristic variables	28345.3264	28368.2288	$F_{4, 151,891} = 30.6946$*
Regional variables	28345.9080	28368.2288	$F_{1, 151,894} = 119.6078$*
Job characteristic variables	28200.6726	28368.2288	$F_{4, 151,891} = 225.6177$*
Participation selection bias corrected estimates (Model 4)			
NURSE dummy (intercept term)	27870.4797	28104.1096	$F_{1, 151,893} = 1273.2739$*
Years of education variables	27846.4546	28104.1096	$F_{2, 151,892} = 702.7060$*
Educational attainment variables	28100.8449	28104.1096	$F_{4, 151,890} = 4.4116$*
Work experience variables	27917.5957	28104.1096	$F_{2, 151,892} = 507.3855$*
Personal characteristic variables	28084.5713	28104.1096	$F_{4, 151,890} = 26.4173$*
Regional variables	28083.3731	28104.1096	$F_{1, 151,893} = 112.1562$*
Job characteristic variables	27940.4598	28104.1096	$F_{4, 151,890} = 222.4084$*
Occupation selection bias corrected estimates (Model 5)			
NURSE dummy (intercept term)	28002.6460	28366.9559	$F_{1, 151,893} = 1976.1035$*
Years of education variables	27990.4351	28366.9559	$F_{2, 151,892} = 1021.6079$*
Educational attainment variables	28361.4809	28366.9559	$F_{4, 151,890} = 7.3303$*
Work experience variables	28118.5200	28366.9559	$F_{2, 151,892} = 671.0066$*
Personal characteristic variables	28345.3121	28366.9559	$F_{4, 151,890} = 28.9949$*
Regional variables	28345.5805	28366.9559	$F_{1, 151,893} = 114.5425$*
Job characteristic variables	28129.3049	28366.9559	$F_{4, 151,890} = 320.8115$*

* Significant at the 5% level

Table 5.12. Results of Chow tests for differences in wage equations between nurses and all other workers

All the results are statistically significant. This means that in all cases the restrictions are too serious to impose. Therefore the structural differences in the wage equations in fact arise from both a difference in the intercept terms and differences in the sets of slope co-efficients tested.

In summary, the wage equation for nurses is different from that for all other workers in terms of the intercept term, and the coefficients on the years of education variables, the educational attainment variables, the work experience variables, the personal characteristic variables, the regional variables and the job characteristic variables. These results are obtained whether or not we correct the wage equation for participation selection bias or occupation selection bias.

The general conclusion is that the two wage equations have different intercepts and different slopes. They are dissimilar regressions.

Given that we now know the wage equations for nurses and all other workers are different we can analyse reasons for the observed wage differential.

5.9. Results of the decomposition analysis

The results are reported in Table 5.13 (they pertain to Models 3, 4 and 5 where separate earnings functions are estimated for nurses and all other workers). The observed difference in mean lnW between nurses and all other workers is 0.3648. We decompose this observed pay differential using equations [5.16] and [5.17] into three main components: due to differences in endowments; due to differences in the returns to endowments; and (where applicable – Models 4 and 5), due to differences in selection bias. The premium to being a nurse is analysed using the characteristics of nurses (using equation [5.16]) and also using the characteristics of all other workers (using equation [5.17]).

In terms of the OLS estimates (Model 3) the differences in endowments is positive (i.e. greater for nurses) and slightly less than the observed difference in mean lnW. The remainder of the difference is explained by a relatively small positive return to endowments (the premium) to being employed as a nurse. These figures imply that nurses earn higher wages than other workers but that this difference is explained primarily by their superior labour market and personal characteristics (the differences in endowments) that influence earnings. For example, as discussed above (Table 5.1) nurses have more years of education than other workers (13.5 versus 13.3 years), a greater proportion of nurses have a nursing qualification

(92% versus 2%), a greater proportion of nurses than all other workers do possess at least some form of educational qualification (99% versus 84%). Nurses also on average tend to have more years of experience with their current employer than other workers (mean 12 years versus mean 10 years). They tend to work longer hours per week (mean 33 versus mean 31 total hours per week), and a larger proportion of nurses play some kind supervisory role in their job (76% of nurses are employed as a supervisor, manager or foreman compared to 24% of all other workers). Also from the OLS estimates we can see that there is a small premium to being employed as a nurse.

	Premium to being a nurse analysed using characteristics of nurses	Premium to being a nurse analysed using characteristics of all other workers
OLS estimates (Model 3)		
Differences in variables (= differences in endowments)	0.3092	0.3134
Premium (= differences in returns to endowments)	0.0556	0.0514
Observed difference in mean lnW	0.3648	0.3648
Participation selection bias corrected estimates (Model 4)		
Differences in variables (= differences in endowments)	0.2882	0.3134
Premium (= differences in returns to endowments)	-0.0942	-0.1194
Differences due to participation selection bias	0.1708	0.1708
Observed difference in mean lnW	0.3648	0.3648
Occupation selection bias corrected estimates (Model 5)		
Differences in variables (= differences in endowments)	0.3602	0.6988
Premium (= differences in returns to endowments)	-0.1312	-0.4698
Differences due to occupation selection bias	0.1358	0.1358
Observed difference in mean lnW	0.3648	0.3648

Table 5.13. Results of decomposition analysis for Models 3, 4 and 5

Decomposition using the participation selection bias corrected estimates (Model 4) yields a different interpretation to explaining wage differentials between nurses and other workers. In this instance the differences in endowments is again positive. This is explained by the labour market and personal factors described above. However, after correcting for participation selection bias there is a negative premium to being employed as a nurse. The differences due to participation selection bias are positive.

Similar results are obtained from the decomposition of wage differentials using the occupation selection bias corrected estimates (Model 5). In this instance the differences in endowments is positive and greater than or equal to the observed difference in mean lnW (depending on whether the premium to being a nurse is analysed using the characteristics of nurses or the characteristics of all other workers). The difference in returns to endowments (the premium to being employed as a nurse) is negative. The differences due to occupation selection bias are positive.

Both these sets of figures from Models 4 and 5 yield a similar interpretation of wage differentials between nurses and all other workers. They imply that while nurses earn higher wages than other workers this pay differential is due exclusively to positive differences in their labour market and personal characteristics. Unlike Model 3, after adjusting for selection bias there is in Models 4 and 5 no evidence of any wage premium to being employed as a nurse. On the contrary, it seems that the average nurse would earn higher wages if paid according to the pay structure of other workers.

5.10. Conclusions

In this chapter we have conducted an earnings function analysis to examine empirically the factors that affect the wages of female nurses working in the NHS. Using the methods outlined in Chapter 4 we have conducted separate analyses that have made no adjustment for selection bias (Model 3), that adjust for potential participation selection bias (Model 4) and that adjust for potential occupation selection bias (Model 5). The data to which the models are applied are taken from the Quarterly Labour Force Survey, a random survey of representative households in Great Britain. The final sample from is taken from the period 1991 to 2000 and consists of 247,774 females aged 18 to 60 years of whom 8,878 are employed as NHS nurses.

There are three important and useful outcomes from the analysis. First, we determined the factors in the wage equation that affect nurses' earnings. We found that important factors positively affecting nurses' earnings are: years of full-time education (there is a concave relationship); possessing a nursing qualification; obtaining a first degree or postgraduate degree; years of work experience (concave relationship); being of non-British nationality; living in the South East of England; [21] and, working as a supervisor, manager or foreman. Factors that have a negative influence of nurses' hourly wages are: possessing A levels as the highest qualification; having no qualifications; having health problems that affect paid work; working longer hours; working at a workplace with 25 or more staff; and, having a non-permanent or temporary job. Ethnic group was found to have a negligible influence on nurses' earnings (the co-efficient on this variable was not statistically significant). An

[21] An interesting finding here is that while the co-efficient on SEAST is positive for nurses it is consistently found to be lower for nurses than for other workers. The interpretation is that while nurses working in the South East of England receive a premium relative to their colleagues working elsewhere this premium is lower for

important point that emphasises the plausibility and robustness of the results is that the estimated co-efficients were generally of the expected sign and order of magnitude and were in the main consistent across the three models. In Model 4 we found that the co-efficient on the participation selection bias correction term was not statistically significant, indicating that for nurses participation selection bias is not significant. In terms of potential occupation selection bias (Model 5), the co-efficient on this variable was found to statistically significant and positive. The interpretation of this co-efficient is ambiguous. Building on these results we then ascertained using multiple Chow tests that the wage equations for nurses and other workers are structurally different, and that this difference lies in terms of both the intercept terms and the slope co-efficients.

From an estimation point of view the second significant finding was the importance of the marginal effects in the wage equation. In the Heckman two-step procedure the full marginal effect on wages of variables that appear as regressors in both the participation/occupation selection equation and the wage equation consists of two components. There is the direct effect on the mean of lnW, which is the co-efficient β in the wage equation. In addition, for independent variables that also appear in the participation/occupation selection equation an indirect effect on lnW will also be exerted through their influence on λ. We estimate the full marginal effects. What is important is that the magnitude, sign and statistical significance of the marginal effects are for many variables different from those of the direct effect given by the relevant co-efficient in the wage equation. This calls into question the conclusions of many earlier earnings function studies that utilise the Heckman two-step approach where the issue of marginal effects is often overlooked.

---

nurses than for other workers. We return to the significance of this finding from a policy perspective in Chapter

The third important outcome from Chapter 5 pertains to the nature and magnitude of the earnings differential between nurses and other workers. We find that nurses in the sample are paid on average higher wages than workers in all other occupations combined. The mean real hourly wages of nurses and all other workers are £7.36 (Std. Dev. £2.96) and £5.49 (Std. Dev. £3.50), respectively. The difference in mean real hourly wages (£1.87 – nurses receive on average 34% higher wages than all other workers) is statistically significant at conventional levels ($p<0.0001$, 95% confidence interval £1.80 to £1.95). Using the algebraic method developed by Oaxaca (1973) we decompose the observed difference in mean ln wages (0.3648) into differences in labour market endowments and differences in the returns to these endowments. The decomposition is informative for the following reason. We wish to compare the earnings of nurses to the earnings of other workers in order to determine whether there is a financial return to being employed as a nurse. One option is to compare the mean earnings of nurses and the mean earnings of all other workers. 'All other workers' however includes both non-manual workers and manual workers some of whom have entirely different years of education, qualifications, and job characteristics to nurses. Thus aside from the potential selection bias problem a raw comparison is not necessarily informative because we may not be comparing like with like. It would be unsurprising to find that a qualified nurse with 15 years experience earns higher hourly wages than cleaner with one year of experience and no post-compulsory schooling and no qualifications, for example. In this chapter the approach adopted is to compare nurses to all other workers in order to utilise the full sample of the available data. The comparison is not problematic because it is possible to disaggregate the observed earnings differential into an endowment component – in which higher earnings are observed due to superior labour market endowments such as schooling, qualifications, etc. – and a premium component which arises due to differences in the returns to

---

7.

endowments. This second effect allows for the possibility that, for example, the impact on earnings of having more experience or having a postgraduate qualification is different for nurses and other workers. In the comparison we wish to control for the differences in endowments. Having then effectively removed the endowment component from the earnings differential we examine the premium. We can conclude that nurses are paid more than other workers if the premium is positive. The implication in this case is that the average nurse would earn lower wages if paid according to the pay structure of other workers. The opposite interpretation is true if the premium is negative. In the selection-bias corrected estimates we find that the endowment effect is positive and the premium is negative. The interpretation is that nurses are paid higher wages than other workers but that this difference is due to differences in nurses' superior labour market and personal characteristics, in particular their greater number of years of education and superior educational attainment, their greater propensity to be employed in a supervisory role and their greater number of years of work experience. The returns to labour market endowments are on average lower for nurses than other workers – the premium is negative. Put another way, after controlling for differences in individual and productive characteristics and selection bias nurses are in fact paid lower wages than other workers.

This is an important and potentially useful finding and has clear implications for reducing the current nursing shortage. However, it should be treated with caution because while the analysis controls for the effect of potential participation selection bias and occupation selection bias the corrections were made individually in separate models. This leads us to Chapter 6 where we construct extended earnings functions for nurses and other workers in Great Britain correcting jointly for *both* participation selection bias *and* occupation selection bias in the same model with in what is known as a 'double selectivity' framework.

CHAPTER 6

DOUBLE SELECTIVITY MODELS OF EARNINGS FOR NURSES

6.1. Introduction

In Chapter 5 we constructed extended earnings functions for nurses in Great Britain using data from the QLFS. The economic theory underpinning the models was based on the work of Mincer (1974), whose specification of the earnings function has been found to predict consistently well earnings patterns and wage differentials across many time frames, settings and geographical areas. We estimated earnings functions for nurses and other workers based on the OLS model and we also augmented our analysis by correcting separately for potential participation selection bias and potential occupation selection bias using the Heckman two-step procedure. This involved including in the wage equations selection bias correction variables (the λ's) that capture the propensity to participate in the labour market and the propensity to be employed as a nurse. In terms of the participation selection bias problem we adjust for the possibility that employees may differ systematically in unobservable characteristics from those who choose not to participate. This would be the case if a non-participator's reservation wage and offered wage were different to those of workers due to their unobserved ability, motivation or personal circumstances. This might arise either because the individual would otherwise earn relatively low wages (they have relatively low offered wages) and therefore the sample of observed wages would be biased upwards, or because individuals who choose not to work might have earned higher wages than those who do choose to work but they have an even higher reservation wage – in which case the sample of observed wages is biased downwards. The second possibility is that occupation selection bias occurs because, for example, individuals self-select into occupations in which they have a comparative advantage in terms of natural ability and motivation. This means that simple comparison of the earnings of nurses and other workers may be a biased estimate of the premium to being employed as a nurse for any given individual. Comparing the actual earnings of nurses and the actual earnings of individuals employed in alternative occupations may overstate or understate the true returns to being employed as a nurse depending on whether an individual employed as a nurse would earn higher or lower wages if employed in another occupation than someone already employed in that occupation and vice versa. In Chapter 5 we found that these sources of selection bias are potentially important factors affecting the earnings of nurses and other workers. The important point is that while we corrected for these two sources of potential bias we made these corrections individually in separate models.

In Chapter 6 we build on this work and construct extended earnings functions for nurses and other workers in Great Britain correcting jointly for *both* participation selection bias *and* occupation selection bias *in the same statistical model*. This is a novel approach which, as we shall see below, has rarely been used in the literature. Following the reasoning given in Chapters 4 and 5 the approach is justified because the regression co-efficients in the wage equation are liable to be biased unless some account is given of the self-selected nature of both the decision to participate and the decision to be employed in a specific occupation. The methodology involves two types of model both of which we shall describe using the umbrella term 'double selectivity model' because they address, in some form or other, both forms of selection bias.

First we estimate two models of earnings based on the bivariate probit selection model (one with and the other without what Greene, 2000, calls censoring, where the observed variables in the bivariate probit model are censored in some way). Essentially this entails including in the wage equations for nurses and other workers two selection bias correction terms that capture the effects of both the participation and the occupation selection decisions.

The second type of model we employ is the multinomial logit selection model. This also entails including in the wage equations for nurses and other workers selection bias correction terms that capture the effects of both the participation and the occupation selection decisions. In the bivariate probit selection models there are two decisions each of which has two alternatives. By contrast, in the multinomial logit model there is a single decision between more than two alternatives. We estimate two models of this type. In the first model – a three-option model (called the trinomial logit selection model) – there are three alternatives: to participate in the labour market as a nurse; to participate in the labour market in an occupation other than nursing; or, to not participate in the labour market at all. We also estimate a four-option model (the quadrinomial logit selection model) for which the alternatives are to participate in the labour market as a nurse, to participate in the labour market in an occupation other than nursing, to be a nurse and to not participate in the labour market, or to be a non-nurse and to not participate in the labour market. The selection bias correction terms included in the wage equations for the two groups of workers (nurses and non-nurses) controls for the self-selected nature of the decision to participate in that particular occupation group relative to the other options. A comparison of the rationale behind each model is presented below.

As noted by Shields and Wheatley Price (1998) jointly controlling for two potential sources of selection bias in the context of a double selectivity problem by any of these methods is rarely investigated in the literature – studies usually concentrate on the more straightforward 'single selectivity' models of Chapter 5, which assume that the regression co-efficients are liable to be biased from only a single source. Further, Moffitt (1999) in his comprehensive survey of 'what labour economists do' finds that over the period 1985 to 1997 econometric methods used in empirical and econometric work in labour economics have utilised hardly ever the techniques on which these selectivity models are based (e.g. bivariate probit and multinomial logit with selection).

We begin this chapter by reviewing the literature on double selectivity models applied to earnings. We focus on extended earnings function studies and then survey empirical applications of this model applied to the labour market for nurses. We then describe the statistical models (bivariate probit selection model, bivariate probit selection model with censoring, trinomial logit selection model, quadrinomial logit selection model) and the structure of the decomposition analysis (section 6.3). Section 6.4 describes the data, the variables and sample characteristics. Sections 6.5 and 6.6 highlight the main results and we present our conclusions in section 6.7.

6.2. Double selectivity earnings functions: the literature

6.2.1. Double selectivity earnings functions

The number of studies using double selectivity earnings functions is currently very small. A search of the *EconLit* database, which is a comprehensive electronic bibliography with coverage of over 400 major economics journals over the period 1969 to 2001, reveals the following limited number of applications: [22]

1. What are the magnitude and causes of male-female earnings differentials in the labour market? (Mohanty, 2001, [BP] Ometto, Hoffman and Correa-Alves, 1999 [ML]) These studies estimate the observed wage differential between males and females in the labour market and correct for selection bias through the worker's decision to participate in the labour market and the employer's decision to hire.
2. Do immigrant workers earn less than native-born workers? (Shields MA. And Wheatley Price, 1998, [BP] Tunali, 1986 [BP]). In these studies, which are conducted separately for England and Turkey, the authors correct for potential selection bias arising from both labour market participation and for the non-reporting of wage information.
3. What are the determinants of workers who are unemployed then re-employed? (Curti, 1998 [BP]) This analysis estimates post-unemployment earnings functions using a double selectivity approach that corrects for both unemployment risk and re-employment probability.
4. What are the characteristics of individuals who become self-employed? (Earle and Sakova, 2000 [ML]) This paper studies the characteristics of the self-employed, employees and the unemployed. Selection-bias-corrected earnings functions are estimated for each group of workers.
5. What are the determinants of earnings of farm-based workers? (Abdulai and Delgado, 1999 [BP]) This study corrects for potential selection bias concerning the decision to participate in the labour market and the decision to undertake farm or non-farm work.

[22] 'BP' denotes an application using the bivariate probit (or independent probit) selection model with or without censoring. 'ML' denotes an application using the multinomial logit selection model.

6. Why do women who work continuously in the labour market earn higher wages than women who work intermittently? (Sorensen, 1993 [BP]). This study estimates earnings functions for intermittent and continuous female workers. Two forms of potential selection bias are controlled for, namely the decision to participate in the labour market and the decision to work intermittently or continuously.
7. What are the factors affecting female earnings? (Zweimuller 1992 [BP]) In this paper the author estimates wage equations for Austrian women which account for potential selection bias due to both non-participation in the labour market and survey non-response.
8. What is the relationship between obesity and earnings? (Pagan and Davila, 1997 [ML]) In this paper the authors utilise a multinomial logit specification to investigate the occupation selection of obese individuals and then estimate earnings functions that account for the occupational attainment of the obese.
9. What is the pay disparity between earnings of females in female-dominated occupations and those in other occupations? (Sorensen, 1989 [BP]) This analysis estimates the earnings differential between women in female-dominated jobs and those with the same characteristics in other occupations. Potential selection bias is controlled for in terms of the individual's decision whether or not to work and for their choice of occupation.
10. What effect does migration have on earnings in low-income countries? (Lanzona, 1995 [ML]) The purpose of this study is to estimate differences in wage equations for individuals who migrate and those who remain in their parental home. Selection bias correction terms are included to control for the self-selected nature of the migration decision.
11. What factors explain the wage differential between public sector and private sector wages? (Gyourko and Tracey, 1986 [ML]) In this paper the authors consider a

multinomial selection model in which worker selects from four labour markets (public sector/union, private sector/union, public sector/non-union, private sector/non-union). Selection-bias corrected wage equations are then estimated for each of the four occupation groups and the wage differentials are computed.

6.2.2. Double selectivity earnings functions for nurses

In addition to the applications discussed above, to date a single study has estimated earnings functions for nurses using a double selectivity model. This study, by Botelho et al. (1998), focuses on the returns to different types of nursing education for US nurses. The study utilises a bivariate framework in that there are two decisions considered (to participate or not, and which path to take to enter the nursing profession). A multinomial logit model is used as well because there are three paths to entering the nursing labour market in the US that are considered in the second decision: by associate degree (AD); by a diploma in nursing (DIP); or, by the bachelor of science in nursing (the baccalaureate degree – BSN). Botelho et al. estimate separate earnings functions for nurses with the three different qualifications to ascertain which avenue into the nursing profession leads to the greatest wage advantage. In their analysis they correct first for participation selection bias and second for potential selection bias arising from the choice of educational credential obtained (BSN, AD or DIP). The authors argue that estimating wage equations with observed data for nurses with each type of qualification by unadjusted OLS potentially gives rise to biased and inconsistent results because individuals who select a particular route into nursing are likely to possess characteristics which predispose them to favour the chosen alternative.

The data used for the analysis is taken from the 1992 National Sample Survey of Registered Nurses which includes 22,147 working nurses, of which 32% possess an associate degree, 41% possess a diploma, and 27% possess a baccalaureate degree. Botelho et al. conduct their analysis of the data in four stages. In the first stage they estimate wage equations for working nurses in each of the three educational credential groups (BSN, AD or DIP). The second stage involves estimation of the participation probit to obtain the selection bias correction term for the likelihood of nursing labour force participation. In the third stage they estimate the selection bias correction term for the credential choice decision using a multinomial logit model. The fourth and final stage of the estimation procedure involves estimation of the augmented (with the λ's) wage equation for each type of nursing qualification by OLS using data on the labour market participants in each education category. A crucial assumption in the analysis is that the two selection bias effects are treated as being independent, which allows the two selection equations to be estimated separately.

The results of the analysis are presented in Table 6.1. From the participation equation results we can see that marital status is a statistically significant determinant of the probability of participation in the labour force: the co-efficient is negative and so married females are less likely to participate than their single counterparts. Being widowed, divorced or separated exerts upward pressure on the likelihood of participation. The co-efficient on having no children is not statistically significantly different from zero at conventional levels. The probability of participating decreases with age and is lower for white nurses.

The results of the multinomial logit model are also presented. These results are interpreted based on the odds ratio. The reference choice for these results is the associate degree (AD), and the co-efficients reflect the effects on the nurse's likelihood of choosing an alternative nursing programme (DIP or BSN).

Variable	Co-efficient		
Participation equation			
Constant	3.5210*		
Married	-0.1403*		
Widowed, divorced or separated	0.1712*		
No children	0.0036		
White ethnic group	-0.1976*		
Age	-0.0498*		
Log-likelihood	-8311.46		
χ^2 statistic	3268.06		
Variable	**Co-efficients**		
Education credential equation [1]	DIP	BSN	
Constant	-2.6426*	-0.1700	
Years of experience	0.1735*	0.0478*	
Employed full-time	0.0019	-0.2905*	
White ethnic group	0.2433*	-0.3890*	
Metropolitan area	0.4066*	0.5845*	
Married	-0.6911*	-0.6383*	
Widowed, divorced or separated	-1.2960*	-1.2314*	
No children	-0.0646	0.2141*	
Log-likelihood	-18,904.96		
χ^2 statistic	11,612.30		
Variable	**Co-efficients**		
Wage equations	AD	DIP	BSN
Intercept	2.6574*	2.2897*	2.7587*
Years of experience	0.0213*	0.0253*	0.0176*
Years of experience squared	-0.0007*	-0.0003*	-0.0004*
Employed full-time	-0.0089	0.0282*	0.0422*
White ethnic group	-0.0844*	-0.0560*	-0.0684*
Metropolitan area	0.1443*	0.1538*	0.1685*
λ_p [2]	-0.1047*	-0.2406*	-0.0608
λ_c [3]	0.0694*	0.1978*	-0.0755*
R^2	0.11	0.06	0.07

[1] The reference group is AD
[2] Participation selection bias correction term
[3] Education credential choice selection bias correction term
* Statistically significant at the 95% level
Source: Botelho et al. (1998)

Table 6.1. Estimates of selection bias-corrected wage equations for AD, DIP and BSN nurses

The results show that nurses with more experience are more likely to have entered the profession via the diploma route. Nurses with a baccalaureate degree are less likely to work full-time than those with AD preparation, while the co-efficient in the diploma equation is insignificant. White nurses are less likely to have chosen a baccalaureate degree than an associate degree, but are more likely to have obtained a diploma in nursing than an associate degree. Holding a nursing job in a metropolitan area increases the likelihood of selecting a diploma or baccalaureate nursing programme over an AD programme. Married, widowed, divorced or separated nurses are more likely to select an associate degree, while nurses without children are more likely to choose the BSN degree relative to AD.

Turning now to the results of the selection-bias-corrected wage equations, we can see that in all three instances the experience-earnings profile is as predicted by the basic Mincerian model, namely an n-shaped parabola. Earnings peak at 15 years, 42 years and 22 years of labour market experience for AD, DIP and BSN nurses, respectively. The co-efficients on the white ethnic group and residing in a metropolitan area dummy variables are of the same sign and order of magnitude. The main differences are in terms of the impact of working full-time and the impact of selection bias. For AD nurses the co-efficient on working full-time is not statistically significantly different from zero. The full-time co-efficient is statistically significant and positive for DIP and BSN nurses. The significance of λ_p and λ_c in all cases save λ_p in the baccalaureate degree equation indicates the presence of potential selection bias effects, which suggests that both the participation decision and the choice of education credential are important in estimating wage equations for US nurses.

Two important limitations of this study are worth noting. First, in terms of the structural model of participation the model is likely to be mis-specified. As we noted in Chapter 4 an important explanatory variable in the participation equation is the amount of property income received, and this is not included in the analysis. Also, the wage equation does not include years of schooling variables, which according to the Mincerian model of earnings are key explanatory variables. Second, an important assumption in this model is that the two sample selection effects are assumed to be independent. It is assumed by the authors that women choose a type of education programme first and then independently choose to participate in the labour market at a later date. The authors acknowledge this as a limitation of their study on the grounds that unobserved factors affecting participation decisions may also be associated with credential choice decisions. For example, holding a baccalaureate degree may be an indicator of those on the fast track in nursing. Nurses who choose this particular option may have clearly defined career goals and exhibit greater attachment to the labour market than those with other credentials.

This is a useful analysis, not least because it is the only study to date to analyse the earnings of nurses using a double selectivity model. However, while revealing of the US nursing labour market, as reasoned above the analysis is probably over-simplified. Further, in terms of British nurses this study is not very illuminating due to substantial differences in labour market structure. One important point gleaned from this study is that from an econometric point of view an important estimation issue in double selectivity models is the dependence of the selection decisions. In other words, in the context of a bivariate probit model to what extent are the two selection equations (e.g. the participation equation and the occupation selection equations) from which the selection bias variables are constructed related? This has implications for the econometric technique used to estimate the model. In the context of binary choices, the two equations may be estimated jointly in a bivariate probit model or separately as independent probits. Fortunately it is possible to test which of the two models is most appropriate given the data. With this context in mind we focus now on the statistical model for the present analysis.

6.3. The statistical models

6.3.1. Bivariate probit selection model

We use a generalised extension of the Heckman two-step procedure (Heckman, 1979; described in detail in Chapter 4) to control for both participation bias and occupation selection bias. The following statistical model is estimated:

Participation equation: $P_i^* = \delta Z_i + V_i$ $(= \ln W_i - \ln W_{ri})$ [6.1]

Occupation selection equation: $Nu_i^* = \gamma z_i + v_i$ $(= \tau_{ni} - \tau_{oi})$ [6.2]

where

$P_i = 1$ if $P^*_i > 0$ (the individual participates in the labour market) [6.3a]

$P_i = 0$ if $P^*_i \leq 0$ (the individual does not participate) [6.3b]

$Nu_i = 1$ if $Nu_i^* > 0$ (the individual is a nurse) [6.3c]

$Nu_i = 0$ if $Nu_i^* \leq 0$ (the individual is not a nurse) [6.3d]

$E(V) = E(v) = 0$, $Var(V) = Var(v) = 1$, $Cov(V, v) = \rho_v$ [6.3e]

P* is an unobserved latent variable reflecting an individual's propensity to participate in the labour market. Z is a vector of regressors influencing labour market participation. Nu* is an unobserved latent variable reflecting whether or not the individual is a nurse. z is a vector of regressors influencing the decision to be a nurse. P and Nu are observed binary variables. δ and γ are vectors of parameters and V and v are error terms. As explained in Chapter 4 P* is determined by the difference between the offered wage ($\ln W_i$) and the reservation wage ($\ln W_{ri}$) – an individual chooses to participate in the labour market if $\ln W_i > \ln W_{ri}$. Nu* is determined by the difference between the present value of lifetime utility if an individual is employed as a nurse (τ_{ni}) and the present value of their lifetime utility if they are employed in another occupation (τ_{oi}). An individual will choose to be employed as a nurse provided $\tau_{ni} > \tau_{oi}$) which will be affected by the individual's productive characteristics and also by their innate ability in different occupations.

For each decision (participate/do not participate, nurse/not nurse) we estimate the selection bias variables – the inverse Mills ratios (λ) – to control for selection bias and include these in the wage equations. The resulting wage equations for nurses and other workers are, respectively:

$$\ln W_{ni} = \beta_n X_{ni} + \beta_{\lambda(p)n}\lambda(p)_{ni} + \beta_{\lambda(nu)n}\lambda(nu)_{ni} + \varepsilon_{ni} \qquad [6.4a]$$

and

$$\ln W_{oi} = \beta_o X_{oi} + \beta_{\lambda(p)o}\lambda(p)_{oi} + \beta_{\lambda(nu)o}\lambda(nu)_{oi} + \varepsilon_{oi} \qquad [6.4b]$$

where

$$E(\ln W_{ni} \mid P^* > 0, Nu^* > 0) = \beta_n X_{ni} + \beta_{\lambda(p)n}\lambda(p)_{ni} + \beta_{\lambda(nu)n}\lambda(nu)_{ni} \qquad [6.5a]$$

$$E(\ln W_{oi} \mid P^* > 0, Nu^* \leq 0) = \beta_o X_{oi} + \beta_{\lambda(p)o}\lambda(p)_{oi} + \beta_{\lambda(nu)o}\lambda(nu)_{oi} \qquad [6.5b]$$

lnW is the natural logarithm of hourly wages. X is a matrix of measurable individual productive characteristics. β is a vector of parameters. $\lambda(p)$ is included as a regressor to reflect the predicted probability of being in paid work given other known characteristics, and $\beta_{\lambda(p)}$ is its coefficient. $\lambda(nu)$ is included as a regressor to reflect the predicted probability of being employed as a nurse given other known characteristics, and $\beta_{\lambda(nu)}$ is its coefficient. ε is an error term and the subscripts distinguish between nurses (n) and all other workers (o).

From the participation equation the inverse Mills ratio for nurses and other workers is estimated as follows:

$$\lambda(p)_i = \frac{\phi(-\delta Z_i / \sigma_V)}{1 - \Phi(-\delta Z_i / \sigma_V)} \qquad [6.6]$$

where ϕ and Φ are, respectively, the standard normal density function and standard normal cumulative distribution function and the p of $\lambda(p)$ denotes an adjustment for participation selection bias. Based on the occupation selection equation we estimate the following inverse Mills ratio for nurses:

$$\lambda(nu)_{ni} = \frac{\phi(\gamma z_i / \sigma_v)}{1 - \Phi(\gamma z_i / \sigma_v)} \qquad [6.7]$$

and the following inverse Mills ratio for all other workers:

$$\lambda(nu)_{oi} = \frac{-\phi(-\gamma z_i / \sigma_v)}{\Phi(-\gamma z_i / \sigma_v)} \qquad [6.8]$$

where the nu of $\lambda(nu)$ denotes an adjustment for occupation selection bias (the decision to become a nurse or not).

The interpretation of the selection bias correction terms $\lambda(p)$ and $\lambda(nu)$ is the same as in the previous two chapters. $\lambda(p)$ may be unambiguously interpreted as measuring the effect of choosing to participate. A positive value of $\beta_{\lambda(p)}$ (its co-efficient) is interpreted to mean that individuals who participate have a higher expected value of lnW than those who choose not to participate, and vice versa. In terms of $\lambda(nu)$ as before the interpretation of the selection bias effect given only the sign of the co-efficient $\beta_{\lambda(nu)}$ on the selection bias correction term $\lambda(nu)$ is ambiguous. The important point is that the inclusion of both variables in the wage equations corrects for the potential participation and occupation selection bias to which the models are liable.

We allow for the fact that the error terms in the two selection equations are correlated (that is, that $\rho_v \neq 0$) by estimating equations [6.1] and [6.2] jointly in a bivariate probit model. This is a testable hypothesis – it is possible to test for correlation among the error terms in the selection equations using the Lagrange multiplier test under the null hypothesis that $\rho_v = 0$. If we reject the null hypothesis it is appropriate to estimate the selection equations jointly. If we fail to reject the null hypothesis it is acceptable to estimate the selection equations separately as independent probits. As explained above researchers commonly assume (but do not test) that $\rho_v = 0$ (see for example Botelho et al., 1998).

In the bivariate probit model first let $q_{iP} = 2P_i - 1$ and $q_{iNu} = 2Nu_i - 1$. Thus $q_{iP} = 1$ if $P_i = 1$ and -1 if $P_i = 0$, and $q_{iNu} = 1$ if $Nu_i = 1$ and -1 if $Nu_i = 0$. The log-likelihood function for the bivariate probit model is:

$$\ln L_{P,Nu} = \sum_{i=1}^{n} \ln\Phi_B(q_{iP}\delta Z, q_{iNu}\gamma z, q_{iP}q_{iNu}\rho) \qquad [6.9]$$

where Φ_B is the bivariate standard normal cumulative density function (see Greene, 2000, for a proof). We maximise this function jointly with respect to δ and γ to estimate the parameters δ and γ.

As in Chapter 5 the wage equations for nurses and other workers are then estimated with the inverse Mills ratios using OLS.

Having estimated the wage equations for nurses and other workers we then decompose the observed nurse/non-nurse wage differential, as in Chapter 5, into three main components: due to differences in endowments; due to differences in the returns to endowments; and due to differences in selection bias. As before we follow the method proposed by Oaxaca (1973), which is to compare the average wages that would be received by nurses and other workers if they were paid according to the same pay structure. Since the OLS estimates of the wage equations pass through the sample mean we have the following output from the wage equations for nurses and other workers, respectively:

$$\ln \overline{W}_n = \hat{\beta}_n \overline{X}_n + \hat{\beta}_{\lambda(p)n} \overline{\lambda}(p)_n + \hat{\beta}_{\lambda(nu)n} \overline{\lambda}(nu)_n \qquad [6.10a]$$

$$\ln \overline{W}_o = \hat{\beta}_o \overline{X}_o + \hat{\beta}_{\lambda(p)o} \overline{\lambda}(p)_o + \hat{\beta}_{\lambda(nu)o} \overline{\lambda}(nu)_o \qquad [6.10b]$$

where the bar indicates a mean value and the hat indicates an estimated value. (Let $\ln \overline{W}$ denote the mean of the natural logarithm of wages.) Subtracting equation [6.10a] from equation [6.10b] we obtain the following wage differentials/decompositions:

$$\ln \overline{W}_n - \ln \overline{W}_o = (\overline{X}_n - \overline{X}_o)\hat{\beta}_o + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}_n + [\hat{\beta}_{\lambda(.)n}\overline{\lambda}(.)_n - \hat{\beta}_{\lambda(.)o}\overline{\lambda}(.)_o] \qquad [6.11a]$$

$$\ln \overline{W}_n - \ln \overline{W}_o = (\overline{X}_n - \overline{X}_o)\hat{\beta}_n + (\hat{\beta}_n - \hat{\beta}_o)\overline{X}_o + [\hat{\beta}_{\lambda(.)n}\overline{\lambda}(.)_n - \hat{\beta}_{\lambda(.)o}\overline{\lambda}(.)_o] \qquad [6.11b]$$

depending on whether the premium to being a nurse is analysed using characteristics of nurses or other workers. The first term on the right hand side of equations [6.11a] and [6.11b] is the contribution to the difference in wages that can be explained by the mean differences in characteristics between nurses and other workers (the difference in the X's). This is referred to as the difference due to endowments, or the difference in variables. The second term provides a measure of the contribution to the differences in returns to characteristics (the β's). This is also called the premium to being a nurse. Note that this premium also captures the difference in intercept terms between the two wage equations. The third term on the right-hand side of equations [6.11a] and [6.11b] describes the joint effects of potential participation and occupation selection bias on wage differentials between nurses and other workers.

The model is estimated using the 'Bivariate Probit Selection Rule' command in LIMDEP version 7.0. (Greene, 1998, p.733-4). The routine for this command does not allow retention of the two jointly estimated λ's which are required for the decomposition analysis. However, it is possible to obtain the means of these variables by running a specially devised programmed sub-routine. The set of LIMDEP commands for doing this is given in Appendix 6.1.

6.3.2. Bivariate probit selection model with censoring

An alternative specification of the bivariate probit selection model is the bivariate probit selection model with censoring. In the bivariate probit setting, and using the notation of the previous section, data on Nu_i might only be observed if $P_i = 1$. For example, whether or not the individual is employed as a nurse ($Nu_i = 0, 1$) is observed only if the individual chooses to participate in the labour market in the first place ($P_i = 1$). Thus there are three types of observation in the sample: $P_i = 0$; $P_i = 1, Nu_i = 0$; and, $P_i = 1, Nu_i = 1$.

The specification of the selection equations is now as follows:

Participation equation: $P_i^* = \delta Z_i + V_i$ [6.12]

Occupation selection equation: $Nu_i^* = \gamma z_i + v_I$ [6.13]

where

$P_i = 1$ if $P^*_i > 0$ (the individual participates in the labour market) [6.14a]

$P_i = 0$ if $P^*_i \leq 0$ (the individual does not participate) [6.14b]

$Nu_i = 1$ if $Nu_i^* > 0$ (the individual is employed as a nurse) [6.14c]

$Nu_i = 0$ if $Nu_i^* \leq 0$ (the individual is employed not as nurse) [6.14d]

$E(V) = E(v) = 0$, $Var(V) = Var(v) = 1$, $Cov(V, v) = \rho_v$ [6.14e]

(Nu_i, z_i) is observed only when $P_i = 1$ [6.14f]

As in the previous section for each decision (participate/do not participate, nurse/not nurse) we estimate the selection bias variables – the inverse Mills ratios (λ) – to control for selection bias and include these in the wage equations. The wage equations for nurses and other workers are specified as before (see equations [6.4a]-[6.5b]).

The log-likelihood function for the bivariate probit model with censoring is:

$$\ln L = \sum_{P=1,Nu=1} \ln \Phi_P(\gamma z_i, \delta Z_i, \rho) + \sum_{P=1,Nu=0} \ln \Phi_P(-\gamma z_i, \delta Z_i, -\rho) - \sum_{P=0} \ln \Phi(-\delta Z_i) \quad [6.15]$$

See Greene (2000) for a proof. We maximise this function jointly with respect to δ and γ to estimate the parameters δ and γ. As before we allow for the fact that the error terms in the two selection equations are correlated (that is, that $\rho_v \neq 0$) by estimating the (nested) selection equations jointly in a bivariate probit model. We test for correlation among the error terms in the selection equations using the Lagrange multiplier test under the null hypothesis that $\rho_v = 0$. If we fail to reject the null hypothesis we estimate the selection equations separately as independent probits.

Having estimated the wage equations for nurses and other workers we then decompose the observed nurse/non-nurse wage differential using the method of Oaxaca (1973) as described above for the uncensored model.

6.3.3. Trinomial logit selection model

An alternative selection model that like the censored bivariate probit selection model allows for three types of observation in the data is the trinomial logit selection model. This differs from the previous models in that we now assume individual i chooses in a single decision to be in one of three possible states of the world: they choose to participate in the labour market as a nurse; they choose to participate in the labour market in an occupation other than nursing; or, they choose not to participate in the labour market at all. Employing a utility-maximising framework, for the i'th individual faced with J choices (in this case J = 3), suppose that the utility of choice j is given by:

$$U_{ji} = \psi_j z_{ji} + \varepsilon_{ji} \qquad [6.16]$$

where z is a vector of characteristics defined as follows:

$$z_{ji} = f(W_{ji}, H_{ji}, C_{ji}, T_{ji}, r_{ji}) \qquad [6.17]$$

where W represents hourly wages, H represents hours worked, C represents personal characteristics, T represents tastes and r is the rate of discount. As in the previous chapter we assume that $W_{ji} = g(X_i, A_{ji})$ where X is a matrix of measurable individual productive characteristics and A represents unmeasurable factors influencing earnings potential for each occupation. ε is an error (disturbance) term. Alternative j is selected if it yields the highest utility. Therefore, the probability that choice j is selected by individual i is:

$$\text{Prob}(U_{ji} > U_{ki}) \text{ for all other choices } k \neq j \qquad [6.18]$$

The model becomes suitable for econometric estimation by making a particular choice for the disturbance term ε. Let D be a polychotomous variable with values 0, 1 and 2 representing the three choices made (0 = choose not to participate in the labour market, 1 = choose to participate in the labour market in an occupation other than nursing, 2 = choose to participate in the labour market as a nurse). McFadden (1973) and Greene (2000) show that if the disturbances are independent and identically distributed with the Weibull distribution

$$F(\varepsilon_{ji}) = \exp(-e^{-\varepsilon_{ji}}),$$

then:

$$\text{Prob}(D_i = j) = \frac{e^{\psi_j z_{ji}}}{\sum_{k=0}^{2} e^{\psi_k z_{ji}}}, \; j = 0, 1, 2 \qquad [6.19]$$

This is the general multinomial logit model applied to a scenario where three alternatives are available. The estimated equations ($\psi_j z_{ij}$) provide a set of probabilities for the choices 0, 1 and 2 for an individual with characteristics z_{ij}. A convenient and useful normalisation of this is to assume that $\psi_0 = 0$. Therefore the probabilities now become:

$$\text{Prob}(D_i = j) = \frac{e^{\psi_j z_{ji}}}{1 + \sum_{k=1}^{2} e^{\psi_k z_{ji}}}, \; j = 1, 2 \qquad [6.20a]$$

$$\text{Prob}(D_i = 0) = \frac{1}{1 + \sum_{k=1}^{2} e^{\psi_k z_{ji}}} \qquad [6.20b]$$

By dividing $\text{Prob}(D_i = j)$ by the probability of $\text{Prob}(D_i = 0)$ we get the odds of the event j occurring. The estimated odds of $D_i = j$ relative to the reference choice $D_i = 0$ are:

$$\frac{\Pr ob(D_i = j)}{\Pr ob(D_i = 0)} = e^{\psi_j z_{ji}} \qquad [6.21]$$

This model therefore allows us to compute the natural logarithm of the odds ratio:

$$\ln\left[\frac{\Pr ob(D_i = j)}{\Pr ob(D_i = 0)}\right] = \psi_j z_{ji} \qquad [6.22]$$

Alternatively we could normalise on any other choice D = k (for $k \neq 0$):

$$\ln\left[\frac{\Pr ob(D_i = j)}{\Pr ob(D_i = k)}\right] = z_{ji}(\psi_j - \psi_k), j = 1, 2, k \neq 0, j \qquad [6.23]$$

This allows us to interpret meaningfully the co-efficients of the estimated equations (the ψ's). In the context of a binary (dummy) independent variable z^1 with co-efficient ψ^1 the relative odds of $D_i = j$ (the odds ratio) which compares those for whom z^1 is present ($z^1 = 1$) with those for whom it is absent ($z^1 = 0$) is:

$$\text{Odds ratio} = \frac{e^{\psi_j{}^0 z_j{}^0 + \psi_j{}^1 + \psi_j{}^2 z_j{}^2 + \ldots}}{e^{\psi_j{}^0 z_j{}^0 + 0 + \psi_j{}^2 z_j{}^2 + \ldots}} = e^{(\psi_j{}^0 z_j{}^0 + \psi_j{}^1 + \psi_j{}^2 z_j{}^2 + \ldots) - (\psi_j{}^0 z_j{}^0 + 0 + \psi_j{}^2 z_j{}^2 + \ldots)} = e^{\psi_j{}^1} \qquad [6.24]$$

and therefore:

$$\ln \text{odds ratio} = \psi_j{}^1 \qquad [6.25]$$

If, for example, z^1 relates to whether or not the individual has a nursing qualification and the model relates to the likelihood of choosing to become a nurse (D = 2) relative to the reference choice of not participating (D = 0), then $e^{\psi_i^1}$ provides the ratio of odds of choosing to be employed as a nurse for those with a nursing qualification to the odds for those without a nursing qualification. If z^1 is a continuous variable (for example, years of education) then ψ_j^1 is instead interpreted as the natural logarithm of the odds ratio relating the odds of choosing to be employed as a nurse to a one-unit increase in z^1.

From the point of view of estimation it is useful that the odds ratio $\left(\frac{\text{Prob}(D_i = j)}{\text{Prob}(D_i = 0)} \text{ or } \frac{\text{Prob}(D_i = j)}{\text{Prob}(D_i = k)}\right)$ does not depend on the other choices. This follows directly from the independence of disturbances in the model. From a behavioural point of view however this is not a very attractive assumption. We return to this problem below.

Based on the multinomial logit model described above, the following selection model is possible (the exposition is based on Lee, 1983, and Greene, 2000). For each of the J decisions for which wage data are available (nurse/non-nurse) we estimate the selection bias variables (the λ's) to control for selection bias and include these in the wage equations. The generalised wage equation is given by:

$$\ln W_j = \beta_j X_{ji} + \beta_{\lambda j}\lambda_{ji} + \eta_{ji}\text{ , } j = 1, 2 \qquad [6.26]$$

where 1 = choose to participate in the labour market in an occupation other than nursing (o, using the terminology of the earlier models), and 2 = choose to participate in the labour market as a nurse (n). The selection bias correction term (λ_j) is estimated as follows:

$$\lambda_{ji} = \frac{\phi[H_j(\psi_j z_{ji})]}{\Phi[H_j(\psi_j z_{ji})]} \qquad [6.27]$$

where H represents the inverse of the standard normal cumulative distribution function evaluated at $Prob(D_i = j)$ $\{H_j = \Phi^{-1}[Prob(D = j)]\}$, ϕ and Φ are, respectively, the standard normal density function and standard normal cumulative distribution function, and the j of λ_j denotes an adjustment for the self-selected nature of choice j. More specifically, the wage equations for nurses and other workers are given by, respectively:

$$\ln W_{2i} = \beta_2 X_{2i} + \beta_{\lambda 2}\lambda_{2i} + \eta_{2i} \qquad [6.28a]$$

and

$$\ln W_{1i} = \beta_1 X_{1i} + \beta_{\lambda 1}\lambda_{1i} + \eta_{1i} \qquad [6.28b]$$

where

$$E(\ln W_{2i} \mid D = 2) = \beta_2 X_{2i} + \beta_{\lambda 2}\lambda_{2i} \qquad [6.29a]$$

$$E(\ln W_{1i} \mid D = 1) = \beta_1 X_{1i} + \beta_{\lambda 1}\lambda_{1i} \qquad [6.29b]$$

In summary, the following two step procedure based on the multinomial logit model is used to adjust for the self-selected nature of the decision to be in state j for j = 0, 1, 2 by (see Greene, 1998, for a more detailed exposition):

Step 1: Estimate the multinomial logit model by maximum likelihood. Retain the co-efficients and the full set of predicted probabilities. Select those observations in the data for which z takes the value in question (z = 1, 2 in this instance). For these observations compute $H_j = \Phi^{-1}[\text{Prob}(D = j)]$ then $\hat{\lambda}_{ji} = \phi[H_j(\psi_j z_{ji})]/\Phi[H_j(\psi_j z_{ji})]$.

Step 2: Estimate β_j and $\beta_{\lambda j}$ by OLS regression of $\ln W_j$ on X_j and $\hat{\lambda}_j$.

The model is estimated using the 'Multinomial Logit Selection Rule' command in LIMDEP version 7.0. (Greene, 1998, p.722-4). The set of LIMDEP commands to do this is quite complicated and is presented in full in Appendix 6.2.

Having estimated the wage equations for nurses and other workers we then decompose the observed nurse/non-nurse wage differential using the method of Oaxaca (1973) as described above.

6.3.4. Quadrinomial logit selection model

An alternative version of the multinomial logit selection model has J = 4 options. In this case like the bivariate probit selection model an individual may be in one of four observable states: to be a non-nurse and to not participate in the labour market (D = 0, the reference choice); to be a nurse and to not participate in the labour market (D = 1); to participate in the labour market in an occupation other than nursing (D = 2); or, to participate in the labour market as a nurse (D = 3). The model proceeds in exactly the same way as before. The normalised probabilities (assuming $\psi_0 = 0$) for choosing the four options are:

$$\text{Prob}(D_i = j) = \frac{e^{\psi_j z_{ji}}}{1 + \sum_{k=1}^{3} e^{\psi_k z_{ji}}}, \; j = 1, 2, 3 \qquad [6.30a]$$

$$\text{Prob}(D_i = 0) = \frac{1}{1 + \sum_{k=1}^{3} e^{\psi_k z_{ji}}} \qquad [6.30b]$$

The co-efficients of the multinomial logit estimation are interpreted as before using the natural logarithm of the odds ratio.

In terms of the wage equations, these are estimated in precisely the same manner as before, though using a slightly different coding to reflect the additional option. The generalised wage equation is given by:

$$\ln W_j = \beta_j X_{ji} + \beta_{\lambda j}\lambda_{ji} + \eta_{ji}, \; j = 2, 3 \qquad [6.31]$$

Note that the selection-bias adjusted wage equations in the two different multinomial logit models will not necessarily be the same. Because the reference choice is different in each case (the reference choice in the quadrinomial logit model is a subset of that used in the trinomial logit model) the computed selection bias variables included in the wage equations (the λ_j's) may well be different in each case.

6.3.5. The independence of irrelevant alternatives

The following is based on Greene (2000) and Freese and Long (2000). We noted in section 6.3.3 that the odds ratios in the multinomial logit model are independent of the other alternatives. This property is convenient as regards estimation but is not a very appealing restriction to place on individuals' behaviour. The property of the multinomial logit model where $\frac{\text{Prob}(D_i = j)}{\text{Prob}(D_i = 0)}$ or $\frac{\text{Prob}(D_i = j)}{\text{Prob}(D_i = k)}$ are independent of the remaining probabilities is called the independence of irrelevant alternatives (IIA). This assumption underpins the statistical estimation of the first stage of the trinomial and quadrinomial logit selection models. The implication is that, for example in the case of the trinomial logit model, the odds ratio between choosing to be employed as a nurse (D = 2) and not participating (D = 0) is independent of the effect of choosing to be employed in an occupation other than nursing (D = 1). This is a testable hypothesis and two procedures have been developed (Hausman and McFadden, 1984, and Small and Hsiao, 1985) to test the validity of the assumption in the data. They work on the principle that if a subset of the choice set (e.g. D = 1) is truly irrelevant then omitting it altogether from the model will not change the remaining parameter estimates systematically: exclusion of the choice from the model will be inefficient but will not lead to inconsistency. Alternatively if the remaining odds ratios are not truly independent of these alternatives then the parameter estimates obtained when these choices are eliminated will be inconsistent.

Hausman and McFadden (1984) proposed a Hausman-type test to test this hypothesis. This test involves the following steps:

1. Estimate the full model with all the J choices included. This is the unrestricted model, with estimates $\hat{\psi}_U$.
2. Estimate a restricted model by eliminating one or more of the outcome categories. This is the restricted model, with estimates $\hat{\psi}_R$.
3. Let $\hat{\psi}_U^*$ be a subset of $\hat{\psi}_U$ after eliminating the co-efficients not estimated in the restricted model.
4. The Hausman test for IIA then has a test statistic given by:

$$H_{IIA} = \left(\hat{\psi}_R - \hat{\psi}_U^*\right)' \left[\hat{V}(\hat{\psi}_R) - \hat{V}\left(\hat{\psi}_U^*\right)\right]^{-1} \left(\hat{\psi}_R - \hat{\psi}_U^*\right) \qquad [6.32]$$

where $\hat{V}(\hat{\psi}_R)$ and $\hat{V}\left(\hat{\psi}_U^*\right)$ are the estimates of the asymptotic covariance matrices. H_{IIA} is distributed as Chi-square with degrees of freedom equal to the number of variables in $\hat{\psi}_R$. Significant values of H_{IIA} indicate that the IIA assumption has been violated.

There is a possibility with this test that restricting the choice set can lead to a singularity. When you drop one or more choices from the analysis in the restricted model one or more of the variables may be constant across the remaining choices. Therefore a case might be induced where there is a regressor which is constant across the choices. One solution is to omit the regressor and re-run the model. This does not introduce a problem of inconsistency or omission of a relevant variable because if the variable is always constant among the remaining choices then variation in it cannot be affecting the choice.

Another problem is that it is common to obtain negative values of the test statistic H_{IIA} (Freese and Long, 2000). H_{IIA} then fails to meet the requirements for the test. Hausman and

McFadden (1984) who developed the test note this possibility and conclude that a negative test result is evidence that IIA has not been violated. However, this is not very satisfactory given the nature and properties of the Chi-squared distribution. A more appropriate interpretation is that the results of the test are ambiguous if H_{IIA} is negative. In these circumstances an alternative option is to use the Small-Hsiao test (Small and Hsiao, 1985).

Small and Hsiao (1985) propose an alternative test based on the same principles. The sample is divided into two random sub-samples of approximately equal sizes. The unrestricted model is estimated on both sub-samples. The weighted average of the co-efficients from the two sub-samples is then defined as follows:

$$\hat{\psi}_U^{S_1S_2} = \left(\frac{1}{\sqrt{2}}\right)\hat{\psi}_U^{S_1} + \left(1 - \frac{1}{\sqrt{2}}\right)\hat{\psi}_U^{S_2} \qquad [6.33]$$

where $\hat{\psi}_U^{S_1}$ is a vector of estimates from the unrestricted model on the first sub-sample and $\hat{\psi}_U^{S_2}$ is its counterpart for the second sub-sample. This has a likelihood given by $L\left(\hat{\psi}_U^{S_1S_2}\right)$. In the second step a restricted sample is created from the second sub-sample by eliminating all cases with the chosen value of the dependent variable. The restricted model is then estimated using the restricted sample yielding the estimates $\hat{\psi}_R^{S_2}$ and the likelihood $L\left(\hat{\psi}_R^{S_2}\right)$. The Small-Hsiao test statistic is the difference between the likelihoods:

$$SH = -2\left[L\left(\hat{\psi}_U^{S_1S_2}\right) - L\left(\hat{\psi}_R^{S_2}\right)\right] \qquad [6.34]$$

SH is distributed as Chi-square with degrees of freedom equal to K+1 where K is the number of independent variables. Significant values of SH indicate that the IIA assumption has been violated.

We test the IIA assumption in the trinomial and quadrinomial logit models using both types of test.

6.3.6. Choice of model

In this section we consider each of the four double selectivity models defined above and distinguish between them in terms of their defining characteristics. The aim is to form an a priori view as to which model is most appropriate in an analysis of nurses' earnings. To aid the exposition we present using option trees in Figures 6.1a to 6.1d a summary of the main features of each of model. Wage equations are estimated for individuals in each state indicated with an asterisk (*).

First we consider differences between the four models in terms of how they model individuals' decisions in the labour market. The appropriate model to use depends on how individuals make participation and occupation selection decisions. In the first two models shown in Figures 6.1a and 6.1b (which utilise the bivariate probit) it is assumed that individuals make two decisions, in each case between a choice of two alternatives. The two decisions are whether or not to participate and which occupation to choose (nurse/non-nurse).

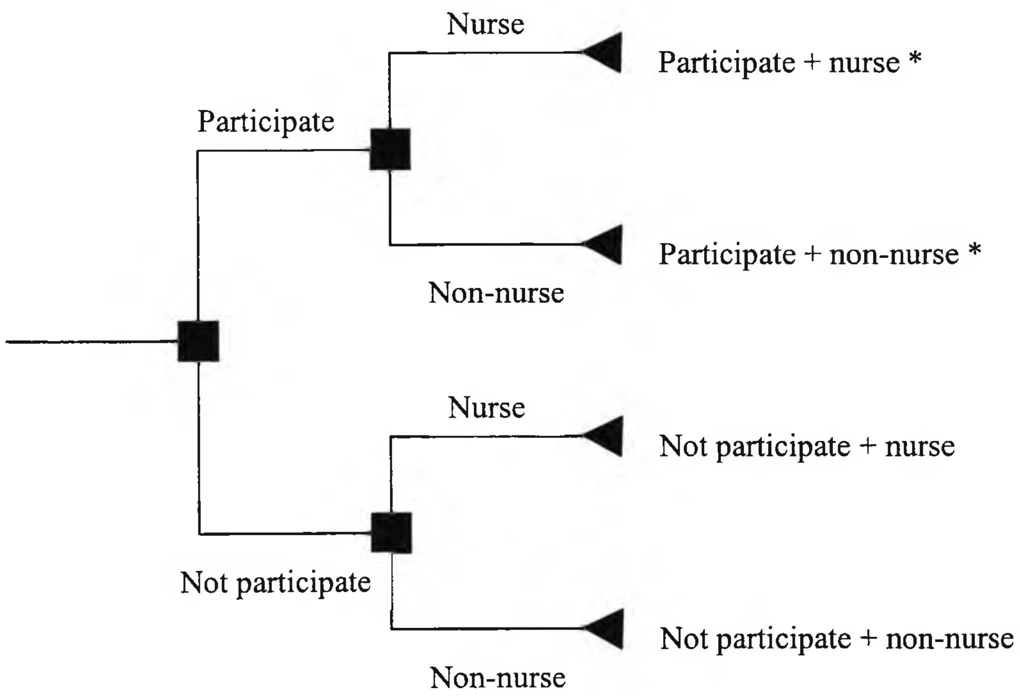


Figure 6.1a. An option tree for the bivariate probit selection model

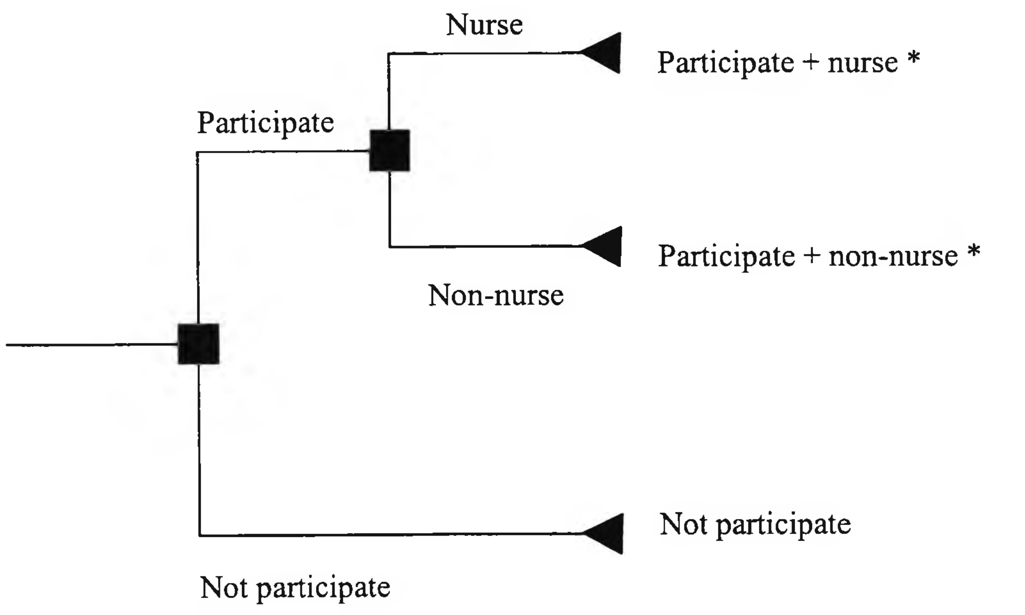


Figure 6.1b. An option tree for the bivariate probit selection model with censoring

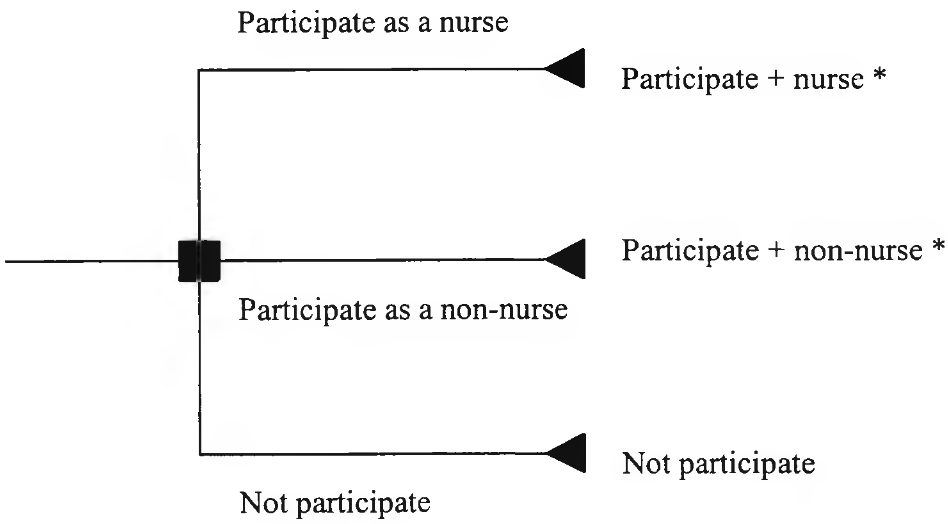


Figure 6.1c. An option tree for the trinomial logit selection model

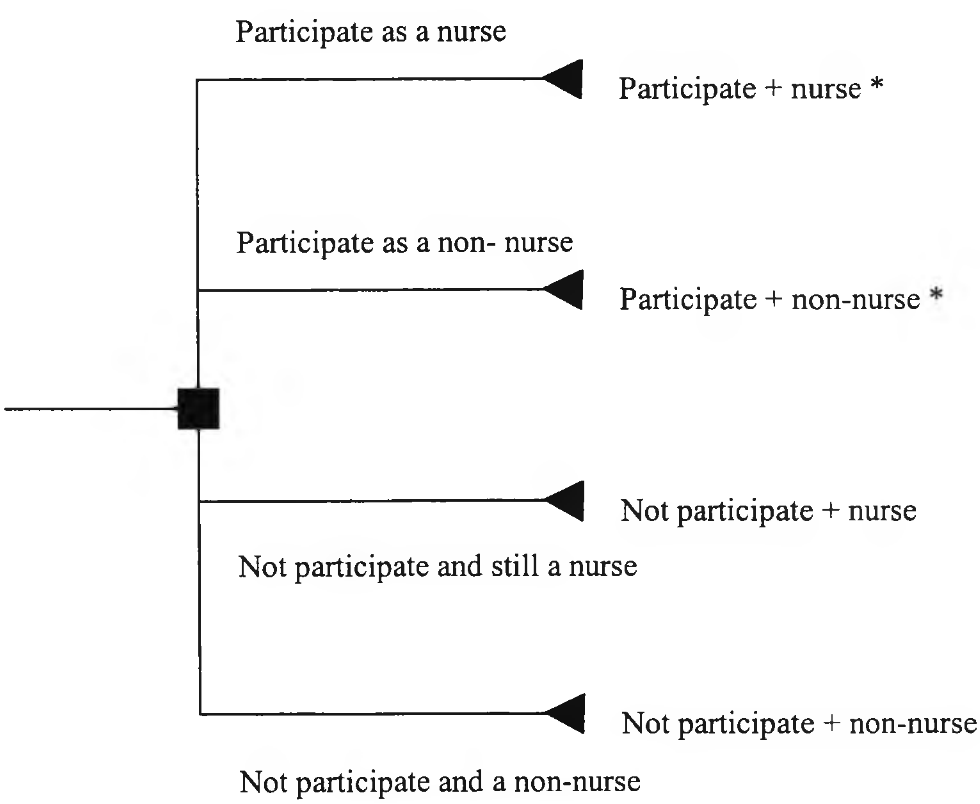


Figure 6.1d. An option tree for the quadrinomial logit selection model

While there are two decisions they may be made jointly by the individual, which implies a joint estimation of the participation and occupation selection equations by bivariate probit. If the two decisions are not made jointly, which is a testable hypothesis, the equations are instead estimated separately as independent probit models. In either case these types of model are appropriate if we believe that individuals decide to participate and then decide in which occupation to work in two (possibly inter-related) decisions. The second set of models shown in Figures 6.1c and 6.1d utilise a multinomial logit framework. This type of model is appropriate if we believe that individuals rather than making two decisions concerning participation and occupation selection instead make a single decision among more than two alternatives (in this case there are three or four alternatives depending on the model used).

Another important distinction should be made between the bivariate probit selection model (Figure 6.1a) and the quadrinomial logit selection model (Figure 6.1d) on the one hand and the bivariate probit selection model with censoring (Figure 6.1b) and the trinomial logit selection model (Figure 6.1c) on the other. The key difference is that in the uncensored bivariate probit model and in the quadrinomial logit model an individual may end up in one of four observable states, reflecting combinations of the participation and occupation selection decisions: participate + nurse; participate + non-nurse; not participation + nurse; and, not participate + non-nurse. There are four terminal nodes to the option trees. In the censored bivariate probit model and in the trinomial logit model however occupation selection is determined given the decision to participate. Therefore, in these two models an individual may end up in one of only three observable states: not participate; participate + nurse; and, participate + nom-nurse. There are only three terminal nodes to these option trees. Which of the two types of model is most appropriate (three or four observable end states) depends fundamentally on how we perceive non-participating individuals. The uncensored bivariate probit model and the quadrinomial logit model allow for non-participating nurses and non-nurses. The assumption is that individuals who do not work may still consider themselves to be 'in' an occupation group in some sense. Reasons for non-participation include: because the individual is unemployed; because the individual is an unpaid family worker (e.g. housewife); or, because the individual is unavailable for or not seeking work because they are sick, disabled or looking after their family. Partly depending on the reasons for non-participation, some non-participating individuals may continue to think of themselves as being in a particular occupation. For example, suppose an individual who was working as a nurse takes a break from employment in order to look after their children. Such an individual might still perceive themselves to be a nurse, in which case one of these two models is more appropriate. The uncensored bivariate probit model and the quadrinomial logit model therefore allow for the fact that individuals might jointly decide to not participate and still maintain a particular occupation decision. An alternative view is that because the individual is not participating (as a nurse, say) then strictly speaking they cannot call themselves a nurse at that point in time and the bivariate probit selection model with censoring or the trinomial logit selection model are more appropriate.

It should be borne in mind that there may be practical difficulties with the data in the uncensored bivariate probit and quadrinomial logit approaches because in datasets that rely on surveys of households occupation data are often collected only for individuals who state that they participate in the labour market. This is clearly problematic for these models, which will require an alternative method for defining the occupation group of non-participators. One possibility is to suppose that holding qualifications relevant to a particular occupation is indicative of being a member of that occupation group. We may therefore define non-participating nurses as individuals who possess a nursing qualification but who do not participate in the labour market at that point in time. Non-participating non-nurses may then be defined as all other non-participators. We follow this coding procedure in the present analysis for estimation of the bivariate probit selection model and the quadrinomial logit selection model. See section 6.4 below for a more detailed explanation. [23]

In terms of the censored model an additional point is that using the bivariate probit framework allows for the possibility that individuals make participation and occupation selection decisions jointly. As noted above this view is testable in practical terms by examining the statistical significance of the correlation co-efficient among the error terms in the selection equations. A priori it is unlikely that there will be such a correlation in this model precisely because the data are censored.

In summary, each model is different and involves an entirely different view of the labour market and in particular how individuals make and perceive their participation and occupation selection decisions. The upshot is that it is difficult to choose between the models in terms of their relevance to individuals' labour market decisions. A case could be made for each model. As noted by Greene (2000) "there is no well-defined testing procedure for discriminating among tree structures...."

An alternative approach is to compare the models on statistical grounds. Rather than examine how realistically each model captures the decision-making process of individuals in the labour market we could instead compare the statistical models in terms of the assumptions and methodologies employed. One potentially important issue is the IIA assumption discussed above in the context of the multinomial logit models. As noted by Greene (2000)

[23] We came across the same issue with Model 4 in Chapter 5 and adopted the same strategy.

"this is not a particularly appealing restriction to place on consumer behaviour." The problem does not apply to the bivariate probit selection models. On the basis of the H_{IIA} and SH tests if the IIA assumption is found not to hold in the trinomial and quadrinomial logit models then there is a case on statistical grounds for preferring the bivariate probit selection models.

6.4. The data and variables

The data used to estimate by regression analysis the four statistical models described in the previous section were taken from the Quarterly Labour Force Survey (QLFS) from Spring 1997 to Autumn 2000. This is the same data source as that used in Chapter 5, the main difference being that here we use only the most recent 15 quarters. The smaller (but still very large) dataset was used to allow for practical computing difficulties in estimating the bivariate probit selection model (this is the largest size of dataset with which it was possible to estimate the bivariate probit model in LIMDEP).

As noted in Chapter 5, the purpose of the QLFS, which is conducted by the Office for National Statistics, is to provide information on the UK labour market that can then be used to develop, manage, evaluate and report on labour market policies. In addition to recording individual, household and family data the QLFS also collects data on economic activity, education and training, health, and, income. The QLFS is representative of the population of the UK; the sample design currently consists of approximately 61,000 randomly selected households in the UK every quarter, representing 0.3% of the population.

Participating nurses are defined in the same way as in Chapter 5 using the Standard Occupational Classification (SOC) code of the occupation in main job of all workers in the sample. SOC code 340 ('registered nurse') was used to define qualified nurses in the analysis. As in Chapter 5 unqualified nurses (nursing auxiliaries and assistants) were not included since their training, qualifications, job specification, work-related skills, and pay are significantly different to that of qualified nurses. Private sector nurses were also excluded from the sample of NHS nurses because they work in a separate labour market with different job specifications and different job characteristics. For the purposes of the analysis both unqualified nurses and private sector nurses were counted as 'non-nurses'. Non-participating nurses are defined as individuals not working in the labour market who possess a nursing qualification. Non-participating non-nurses are defined as all other non-participators.

The final sample comprises 125,778 females aged 18 to 60 years of whom 80,694 participate in the labour market. For the bivariate probit selection model and the quadrinomial logit selection model the sample is divided into four sub-samples: participating nurses (n = 3,461); participating non-nurses (n = 77,233); non-participating nurses (n = 1,050); and, non-participating non-nurses (44,034). In the bivariate probit selection model with censoring and the trinomial logit selection model there are three sub-samples: participating nurses (n = 3,461); participating non-nurses (n = 77,233); and, non-participators (n = 45,084). In all four models wage information is observed only if P = 1. The proportion of workers in the sample (64%) and the proportion of workers who are employed as nurses (4%) are comparable to those in the larger dataset (of which the data used here are a subset) used in Chapter 5.

The dependent variable in the wage equations (LNWAGE) is the natural logarithm of the hourly wage measured in constant December 1992 UK£. Hourly wages were computed as usual gross weekly pay divided by total usual hours worked per week. Hourly wages were used (instead of daily, weekly, monthly or annual wages) to allow for the effect of total hours worked on total wages. Wages were converted to constant December 1992 prices using the monthly retail prices index (Office for National Statistics, selected years).

Mean values for the variables used in the analysis and their definitions are presented in Table 6.2. The full set of descriptive statistics of variables used in the wage equations, including the time trend variables, are presented in Appendix 6.3. Independent variables included in the wage equations are years of education variables (YED, YED2) and work experience variables (EXP, EXP2) – the inclusion of both of which is consistent with the Mincerian model of earnings, educational attainment variables (PGDEG, DEG, ALEVEL, NOQUAL), personal characteristic variables (DISABLE, ETHNIC, NONBRIT and ETHNBRIT), a regional dummy variable for whether or not the individual lives in the South East of England (SEAST), and job characteristic variables (HOURSPW, MANAGE, NWORKERS and TEMP). Additionally 14 quarterly dummy variables are included to allow for the heterogonous nature of the time dimension in the data (the time trend variables). Independent variables included in the participation equation and the occupation selection equation are years of education variables, educational attainment variables, personal characteristic variables, regional variables, time trend variables, age variables (AGE, AGE2), family variables (PCHILD, COHABIT, MARRIED), and property income variables (PENSION, NONLABY).

As in Chapter 5 we base identification of the wage equation on exclusion of the property income variables (PENSION and NONLABY). A priori it is posited that these variables will be good (i.e. statistically significant) predictors of participation and occupation selection but will not be associated with actual observed wages (the dependent variable in the wage equation) when controlling for other covariates.

Nurses in the data are paid on average higher wages than workers in other occupations. In the sample the mean real hourly wages of nurses and all other workers are £7.42 (Std. Dev. £2.37) and £5.59 (Std. Dev. £3.44), respectively (data not shown). The difference in mean real hourly wages (£1.83 – nurses receive on average 33% higher wages than other workers) is statistically significantly different from zero ($p<0.001$, 95% confidence interval £1.75 to £1.91). This wage differential is comparable to that observed in the larger dataset utilised in Chapter 5 (£1.87).

Most of the differences between nurses and other workers in the means of the variables included in the wage equations are statistically significant. Nurses have slightly more years of full-time education (mean 13.7 versus 13.4 years). A greater proportion of nurses have a nursing qualification (92% versus 2%) than the rest of the working population though nurses are generally less well educated at the top end of the educational attainment spectrum (in terms of postgraduate and first degrees). A greater proportion of nurses do possess at least some form of educational qualification, however (99% of nurses versus 87% of other workers possess an educational qualification). Nurses on average tend to have more years of work experience than other workers (mean 10 years versus mean 8 years). In terms of the personal characteristics variables there are small differences between nurses and other workers in terms of the prevalence of health problems affecting paid work of workers (6% for nurses versus 7% for other workers). Slightly more nurses are from non-white ethnic groups (5% versus 4%) and are non-British (6% versus 4%). Fewer nurses live in the South East of England (26% versus 30%) which might be important because individuals living in this region receive extra wage payments in terms of a London weighting allowance to help cover the increased costs of living. Nurses tend to work longer hours than other workers (mean 34 versus mean 31 total hours per week), and a larger proportion of nurses play some kind supervisory role in their job (74% of nurses are employed as a supervisor, manager or foreman compared to 25% of all other workers). Nurses on average tend to work in larger establishments in terms of numbers of workers employed (83% of nurses work in establishments with 25 or more total workers, compared with 63% of other workers), and a smaller proportion of nurses are employed on temporary contracts (6% versus 8%). In terms of the time trend variables (data not shown – see Appendix 6.3), each quarter in the dataset contains between 5% and 10% of all observations. The proportions are similar across nurses and other workers.

	Entire sample [1]		Nurses [2]		All other workers [2]		Definition
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	
LNWAGE*			1.9471	0.3790	1.5882	0.5125	LN hourly wage
NURSE	0.0359	0.1860					Employed as a nurse=1, 0 otherwise
PART	0.6416	0.4795					Participate in the labour market=1, 0 otherwise
Years of education variables							
YED*	13.2325	2.7559	13.6726	2.1281	13.4141	2.5825	Years of full-time education
YED2*	182.6930	82.2442	191.4690	65.2364	186.6080	78.1909	Years of full-time education squared
Educational attainment variables							
NURSEQUA*	0.0485	0.2149	0.9151	0.2788	0.0244	0.1543	Has a nursing qualification=1, 0 otherwise
PGDEG*	0.0304	0.1717	0.0173	0.1305	0.0411	0.1984	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG*	0.0802	0.2717	0.0878	0.2831	0.1005	0.3007	Highest qualification is a first degree=1, 0 otherwise
ALEVEL*	0.0720	0.2585	0.0090	0.0942	0.0730	0.2602	Highest qualification is A level=1, 0 otherwise
NOQUAL*	0.1983	0.3987	0.0035	0.0588	0.1318	0.3383	Has no qualifications=1, 0 otherwise
Work experience variables							
EXP*			10.3216	8.1480	7.8358	6.9218	Years of experience with current employer
EXP2*			172.9050	242.0870	109.3110	185.9030	Years of experience with current employer squared
Personal characteristic variables							
DISABLE[#]	0.1568	0.3636	0.0601	0.2377	0.0680	0.2517	Health problems affect paid work =1, 0 otherwise
ETHNIC*	0.0594	0.2364	0.0511	0.2203	0.0383	0.1919	Non-white ethnic group=1, 0 otherwise
NONBRIT*	0.0561	0.2300	0.0618	0.2409	0.0393	0.1944	Non-British nationality=1, 0 otherwise

ETHNBRIT*	0.0191	0.1368	0.0165	0.1273	0.0093	0.0961	Non-white and non-British=1, 0 otherwise
Regional variables							
SEAST*	0.2922	0.4548	0.2641	0.4409	0.2997	0.4581	Lives in the South East of England=1, 0 otherwise
Job characteristic variables							
HOURSPW*			33.6053	10.5120	31.4290	12.9791	Total usual hours worked per week
MANAGE*			0.7437	0.4366	0.2479	0.4318	Employed as a supervisor, manager or foreman=1, 0 otherwise
NWORKERS*			0.8310	0.3748	0.6298	0.4829	25+ workers at workplace=1, 0 otherwise
TEMP*			0.0569	0.2317	0.0754	0.2641	Job is non-permanent or temporary=1, 0 otherwise
Age variables							
AGE	38.6731	11.4875					Years of age
AGE2	1627.5700	900.8340					Years of age squared
Family variables							
PCHILD	0.1324	0.3390					Age 20-29 years and cohabiting or age 25-34 years and married=1, 0 otherwise
COHABIT	0.1000	0.3000					Cohabiting (living as a couple but not married)=1, 0 otherwise
MARRIED	0.5753	0.4943					Married=1, 0 otherwise
Property income variables							
PENSION	0.0186	0.1351					Receives an occupational pension=1, 0 otherwise
NONLABY	121.1280	1059.3200					Non-labour income
N	125,778		3,461		77,233		

[1] Summary statistics for variables included in the participation and occupation selection equations. Includes workers and non-workers.

[2] Summary statistics for variables included in the wage equations. Includes workers only.

* Difference in mean values between nurses and all other workers significant at the 5% level.

[#] Difference in mean values between nurses and all other workers significant at the 10% level.

Table 6.2. Sample means and standard deviations

6.5. Results of the statistical models

For comparison with the results of this chapter we present in Appendices 6.4 to 6.9 the results of six statistical models based on the formulations used in the previous chapter. The main difference between these results and those of Chapter 5 are that those presented here are estimated on the sub-sample of the data from Spring 1997 (quarter 17) to Autumn 2000 (quarter 31). Appendix 6.4 presents the results of a simple OLS extended earnings function including a dummy variable for whether or not an individual is employed as a nurse (Model 1 in Chapter 5). Appendix 6.5 is the participation selection bias model described in Chapter 4 (Model 2). Appendix 6.6 provides separate OLS earnings functions for each occupation group with no adjustment for selection bias (Model 3). Appendix 6.7 presents separate estimates of wage equations for nurses and other workers. Adjustments are also made to the separate wage equations for participation selection bias using the Heckman two-step procedure (Model 4). In Appendix 6.8 we present the results of the occupation selection bias model described in Chapter 4. We estimate wage equations separately for nurses and other workers, with adjustments to the occupation-specific wage equations for occupation selection bias (Model 5). Finally, we present in Appendix 6.9 the results of an alternative occupation selection bias model (version 2), estimating wage equations separately for nurses and other workers with adjustments to the occupation-specific wage equations for occupation selection bias as before. The difference between this and the previous version is that the occupation selection equation in this model is estimated on the entire sample (rather than on workers only).

For each set of results presented in this section the co-efficients for the time trend variables are not shown. See Appendix 6.10 for the full set of results including those pertaining to the time trend variables.

6.5.1. Bivariate probit selection model

The results of the jointly estimated participation equations and occupation selection equations for the bivariate probit selection model are presented in Table 6.3. Participation is positively related to cohabiting and being married, obtaining a postgraduate or first degree and living in the South East of England. It is negatively related to having health problems that affect paid work, being non-white and/or non-British, having children, and having A levels as the highest qualification, or having no educational qualifications at all. The co-efficients on the property income variables are also statistically significant and of the expected sign. There are concavities in the relationship between age and years of education and propensity to participate in the labour market. In terms of the occupation selection decision, factors positively affecting the decision to become a nurse are having a nursing qualification, being of non-white ethnic origin or being non-British, having children and receiving a pension. Factors negatively affecting the propensity to be a nurse are being married, having a postgraduate qualification and having no educational qualifications.

An important feature of this model is that the participation and occupation selection equations are estimated in a single step to allow for a possible correlation between the random error terms in the two binary choices. The correlation co-efficient measures the correlation between the disturbances in the participation and occupation selection equations. That is, ρ_v measures the correlation between the outcomes (the propensity to participate and the propensity to be a nurse) after the influence of the included factors (the independent variables in the two equations) is accounted for. In the analysis we find that ρ_v is statistically significantly different from zero and negative (-0.1386). This means that unobserved factors affecting labour market participation such as innate ability, motivation or personal circumstances are also associated (negatively) with the decision to become a nurse. The implication as far as model estimation goes is that the two selection equations should be estimated jointly as is the case here and not as two separate probit models. For comparison in Appendix 6.11 we present the results of the participation and occupation selection equations estimated as independent probits. We also present the resulting wage equations for nurses and other workers with the two independently-estimated selection bias correction terms included.

	Participation equation		Occupation selection equation	
	δ [1]	Std.Err.	γ [2]	Std.Err.
Constant	-2.2921*	0.0749	-2.6670*	0.3572
NURSEQUA			3.3397*	0.0323
Age variables				
AGE	0.0739*	0.0025	-0.0018	0.0101
AGE2	-0.0008*	0.00003	-0.00002	0.0001
Personal characteristics				
DISABLE	-1.0734*	0.0111	0.0312	0.0556
ETHNIC	-0.4530*	0.0199	0.1673*	0.0648
NONBRIT	-0.3575*	0.0201	0.2883*	0.0532
ETHNBRIT	-0.0055	0.0391	-0.1100	0.1184
Family variables				
PCHILD	-0.0799*	0.0136	0.1487*	0.0435
COHABIT	0.2449*	0.0146	-0.0368	0.0503
MARRIED	0.0541*	0.0108	-0.0807*	0.0341
Property income variables				
PENSION	-0.6753*	0.0298	0.3288*	0.1105
NONLABY	-0.00004*	0.000002	-0.00003	0.00002
Years of education variables				
YED	0.2015*	0.0072	0.0116	0.0344
YED2	-0.0069*	0.0002	-0.0006	0.0011
Educational attainment variables				
PGDEG	0.5524*	0.0271	-0.7776*	0.0706
DEG	0.3340*	0.0171	-0.0415	0.0401
ALEVEL	-0.0913*	0.0155	0.0312	0.0639
NOQUAL	-0.6185*	0.0108	-0.5505*	0.0887
Regional variables				
SEAST	0.0155#	0.0086	-0.0372	0.0286
ρ_v		-0.1386*		
Log likelihood function		-75,693.25		
N		125,778		

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[2] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0)
* Significant at the 5% level
# Significant at the 10% level

Table 6.3. Results of participation and occupation selection equations estimated jointly by bivariate probit

We turn now to the results of the selection bias corrected wage equation estimates for nurses and all other workers, presented in Table 6.4. Identification of the wage equations is achieved through the omission of the property income variables (see Appendix 6.12).

The co-efficients are of the expected sign and order of magnitude, and are consistent with the results of Chapter 5. They are consistent with the Mincerian model of earnings, though there are differences between the co-efficients for nurses and all other workers. In terms of the years of education variables the earnings-years of education profile for both nurses and all other workers is n-shaped. For nurses the maximum earnings occur after 16 years of education, and for all other workers the maximum occurs at 20 years. The co-efficients on the work experience variables also indicate a concavity in the experience-earnings profile. In this case the relationship between work experience and earnings is also n-shaped with earnings maximised at 26 and 31 years of work experience for nurses and all other workers, respectively. In terms of the educational attainment variables the transformed co-efficients on NURSEQUA suggests that the wage premium to obtaining a nursing qualification is greater for nurses than other workers (30% versus 13%). [24] The benefits in terms of increased wages to having a postgraduate degree or first degree are, while positive, lower for nurses than other workers (returns to a postgraduate degree are 20% for nurses and 44% for workers in other occupations, and for a first degree they are 9% and 33%, respectively).

White non-British nurses earn 10% higher wages than white British nurses, with a slightly smaller premium for white non-British workers in other occupations (5%). For individuals employed as nurses the other personal characteristic variables are not statistically significantly different from zero. Non-white British workers employed in occupations other than nursing receive on average a 3% premium, while for non-white non-British workers the effect is negative (-10%). Regional effects are less pronounced for nurses than all other

[24] As noted in Chapter 5 Halvorsen and Palmqvist (1980) show that to obtain the relative effect g on wages W of a dummy variable given its co-efficient β in a wage equation with dependent variable lnW we must calculate $g = e^{\beta} - 1$, i.e. take the antilog of β and subtract one. The percentage effect is therefore equal to $100*g = (e^{\beta} - 1)*100$.)

workers. Nurses residing in the South East of England earn on average 7% higher wages than other nurses, while for all other workers the average effect on lnW is 18%. In terms of the job characteristic variables, nurses' hourly wages are inversely related to the number of hours worked, but there is an opposite effect, on average, for all other workers. There is a wage premium to being employed in a managerial position for both nurses and all other workers (5% and 15%, respectively). For nurses, being employed in a relatively large workplace or being employed on a temporary contract are negatively related to earnings (for other workers the effect on mean wages of being employed in a workplace with 25 or more workers is positive).

For non-nurses the co-efficients on the time trend variables are as expected variables (data not shown – see Appendix 6.10): they are generally statistically significant; have the expected sign (positive); and, are generally of the expected rank order of magnitude (i.e. generally the co-efficients increase with time, albeit in a non-uniform manner). For nurses the picture is somewhat different, the main point being that only after quarter 27 (Autumn 1999) are the co-efficients statistically significant. What this seems to indicate is that after controlling for measurable individual productive characteristics and selection bias real wage increases over time for nurses between Spring 1997 and Autumn 1999 (quarters 17 to 27) were not significant.

For nurses neither of the co-efficients on the selection bias variables is statistically significantly different from zero. This may be interpreted to mean that selection bias is not significant for the group. For non-nurses the co-efficient on $\lambda(p)$ is negative which means that non-nurses who participate will earn lower expected wages than (the same) individuals who do not participate would earn if they chose to participate. The co-efficient on $\lambda(nu)$ for non-nurses is not statistically significant.

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.8564*	0.2084	0.2250*	0.0398
Years of education variables				
YED	0.1052*	0.0247	0.0997*	0.0052
YED2	-0.0033*	0.0008	-0.0025*	0.0002
Educational attainment variables				
NURSEQUA	0.2609*	0.1041	0.1208#	0.0673
PGDEG	0.1863*	0.0552	0.3661*	0.0092
DEG	0.08718*	0.0241	0.2826*	0.0064
ALEVEL	-0.0584	0.0677	0.0467*	0.0063
NOQUAL	-0.0844	0.1461	-0.1123*	0.0059
Work experience variables				
EXP	0.0208*	0.0024	0.0308*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristics variables				
DISABLE	0.0238	0.0505	-0.0112	0.0088
ETHNIC	0.0324	0.0376	0.0259*	0.0095
NONBRIT	0.0843*	0.0336	0.0683*	0.0093
ETHNBRIT	0.0029	0.0644	-0.1075*	0.0205
Regional variables				
SEAST	0.0653*	0.0143	0.1632*	0.0034
Job characteristics variables				
HOURSPW	-0.0037*	0.0006	0.0018*	0.0001
MANAGE	0.0526*	0.0147	0.1435*	0.0038
NWORKERS	-0.0377*	0.0164	0.1341*	0.0033
TEMP	-0.0749*	0.0273	0.0073	0.0061
Selection bias variables				
$\lambda(p)$	-0.1193	0.0746	-0.0532*	0.0047
$\lambda(nu)$	0.0285	0.0419	0.0246	0.0555
Adjusted R^2	0.1322		0.3023	
Model test	$F(34, 3{,}426) = 16.50$; $p = 0.0000$		$F(34, 77{,}198) = 985.22$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table 6.4. Results of selection-bias corrected wage equation estimated in the bivariate probit selection model

Notice that the relative importance of the selection bias variables changes depending on the statistical model used. Most noticeably, in the rejected double selectivity independent probit model (Appendix 6.11 Table A6.11.2) both the selection bias variables for nurses are statistically significantly different from zero, which is not the case in the bivariate probit model. This would seem to call into question the results of previous studies where the independence of the two selection rules is simply assumed and not tested (see for example Shields and Wheatley Price, 1998, and Botelho et al., 1998).

It is also interesting to note, however, that the statistical significance and sign of the selection bias variables in the bivariate probit model is the same as those in the separately-estimated participation-selection-bias-corrected wage equations (Appendix 6.7, Table A.6.6.2) and occupation-selection-bias-corrected wage equations (Appendix 6.8, Table A6.8.2 and Appendix 6.9, Table A6.9.2) which are estimated using the Heckman two-step method.

6.5.2. Bivariate probit selection model with censoring

The key feature of the bivariate probit selection model with censoring is that data on Nu_i is observed if $P_i = 1$. That is, whether or not the individual is employed as a nurse ($Nu_i = 0, 1$) is observed only if the individual chooses to participate in the labour market in the first place ($P_i = 1$). The results of the jointly estimated participation equations and occupation selection equations are presented in Table A6.13.1 in Appendix 6.13. Note that the correlation co-efficient ρ_v measuring the correlation between the disturbances in the participation and occupation selection equations is not statistically significantly different from zero. The value of ρ_v is –0.2354, and its standard error is 0.3633. A priori this finding is unsurprising because occupation selection data are only observed when the individual chooses to participate and so the observations in the data used to calculate ρ_v are censored by model structure to include only those who participate. The implication however is that it is appropriate to estimate the two selection equations separately as independent probit models (called the independent probit selection model with censoring). It is to these results that we now turn.

The results of the participation and occupation selection equations estimated in the independent probit model with censoring are presented in Table 6.5.

	Participation equation		Occupation selection equation	
	δ [1]	Std.Err.	γ [2]	Std.Err.
Constant	-2.2896*	0.0897	-2.6536*	0.3938
NURSEQUA			3.0340*	0.0306
Age variables				
AGE	0.0739*	0.0025	0.0120	0.0102
AGE2	-0.0008*	0.0000	-0.0002#	0.0001
Personal characteristics variables				
DISABLE	-1.0733*	0.0111	-0.2198*	0.0581
ETHNIC	-0.4529*	0.0197	0.2253*	0.0716
NONBRIT	-0.3575*	0.0202	0.3365*	0.0640
ETHNBRIT	-0.0060	0.0395	-0.0487	0.1387
Family variables				
PCHILD	-0.0798*	0.0137	0.1351*	0.0449
COHABIT	0.2450*	0.0145	-0.0472	0.0495
MARRIED	0.0540*	0.0106	-0.1114*	0.0366
Property income variables				
PENSION	-0.6746*	0.0298	-0.0246	0.1386
NONLABY	-0.00004*	0.000005	-0.00003#	0.00002
Years of education				
YED	0.2012*	0.0100	0.0087	0.0451
YED2	-0.0069*	0.0003	-0.0006	0.0015
Educational attainment variables				
PGDEG	0.5524*	0.0281	-0.7110*	0.0897
DEG	0.3341*	0.0175	-0.0482	0.0481
ALEVEL	-0.0910*	0.0156	0.0538	0.0684
NOQUAL	-0.6186*	0.0107	-0.3629*	0.0923
Regional variables				
SEAST	0.0155#	0.0087	-0.0550#	0.0307
Log likelihood function	-70,023.34		-5,120.661	
Restricted log likelihood	-82,072.15		-14,284.75	
Model test	$\chi^2 = 24{,}097.62$; df = 32; sig. = 0.0000		$\chi^2 = 18{,}328.18$; df = 33; sig. = 0.0000	
N	125,778		80,694	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[2] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0). This only observed when PART = 1.
* Significant at the 5% level
# Significant at the 10% level

Table 6.5. Results of participation and occupation selection equations estimated by independent probit with censoring

As in the previous model participation is positively related to cohabiting and being married, obtaining a postgraduate or first degree and living in the South East of England. It is negatively related to receiving a pension, increases in non-labour income, experiencing health problems that affect paid work, being non-white and/or non-British, having children, and having A levels as the highest qualification, or having no educational qualifications at all. There are concavities in the relationship between age and years of education and propensity to participate in the labour market. In terms of the occupation selection decision, factors positively affecting the decision to become a nurse given the decision to participate are having a nursing qualification, being of non-white ethnic origin or being non-British, and having children. Factors negatively affecting the propensity to be a nurse are being married, experiencing health problems that affect paid work, having a postgraduate qualification and having no educational qualifications at all.

The results of the selection bias corrected wage equation estimates for nurses and all other workers in the independent probit selection model with censoring are presented in Table 6.6. As before the co-efficients are consistent with the Mincerian model of earnings, though there are differences between the co-efficients for nurses and all other workers. In terms of the years of education variables for nurses the maximum earnings occur after 16 years of education, and for all other workers the maximum occurs at 37 years. The co-efficients on the work experience variables also indicate a concavity in the experience-earnings profile. In this earnings are maximised at 24 and 30 years of work experience for nurses and all other workers, respectively. The benefits in terms of increased wages to having a postgraduate degree are similar for both occupation groups, indicating a premium of 25%. For first degrees the effect is much smaller for nurses than for other workers (7% versus 21%).

In terms of the personal characteristic variables the co-efficient on DISABLE in the wage equation for nurses indicates that, given their decision to participate, nurses with health problems that affect paid work will on average earn 15% higher wages than other nurses. For other workers the effect is more pronounced at +33%. White non-British nurses earn 10% higher wages than white British nurses, with a larger premium for white non-British workers in other occupations (17%).

Regional effects are less pronounced for nurses than all other workers. Nurses residing in the South East of England earn on average 7% higher wages than other nurses, while for other workers the average effect on lnW is 17%.

In terms of the job characteristic variables, as in the previous models nurses' hourly wages are inversely related to the number of hours worked, but there is an opposite effect, on average, for all other workers. There is a wage premium to being employed in a managerial position for both nurses and all other workers (6% and 15%, respectively). For nurses, being employed in a relatively large workplace or being employed on a temporary contract are negatively related to earnings, while for all other workers on average these effects are positive.

For nurses the co-efficient on the participation selection bias variables is statistically significantly different from zero and negative, which is interpreted to mean that nurses who participate will earn lower expected wages than (the same) individuals who do not participate would earn if they chose to participate. For non-nurses the co-efficient on $\lambda(p)$ is also negative and has the same interpretation. Occupation selection bias is not statistically significant for nurses. For other workers the co-efficient on the occupation selection bias correction term is statistically significant though, as noted previously, its interpretation is ambiguous in this context.

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	1.5009*	0.3973	0.7690*	0.0420
Years of education variables				
YED	0.0886*	0.0259	0.0532*	0.0052
YED2	-0.0027*	0.0009	-0.0007*	0.0002
Educational attainment variables				
NURSEQUA	-0.1074	0.2319	0.0449	0.0595
PGDEG	0.2239*	0.0659	0.2246*	0.0101
DEG	0.0671*	0.0255	0.1923*	0.0069
ALEVEL	-0.0641	0.0680	0.0817*	0.0063
NOQUAL	-0.0790	0.1169	0.0358*	0.0071
Work experience variables				
EXP	0.0204*	0.0024	0.0267*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	0.1396*	0.0636	0.2883*	0.0121
ETHNIC	0.0488	0.0389	0.1468*	0.0101
NONBRIT	0.0858*	0.0363	0.1533*	0.0096
ETHNBRIT	0.0101	0.0644	-0.0935*	0.0203
Regional variables				
SEAST	0.0699*	0.0147	0.1563*	0.0034
Job characteristic variables				
HOURSPW	-0.0038*	0.0006	0.0019*	0.0001
MANAGE	0.0516*	0.0147	0.1386*	0.0038
NWORKERS	-0.0375*	0.0165	0.1337*	0.0033
TEMP	-0.0755*	0.0272	0.0100#	0.0060
Selection bias variables				
λ(p)	-0.2811*	0.0866	-0.5741*	0.0162
λ(nu)	-0.1371	0.0979	0.1072#	0.0601
Adjusted R^2	0.1333		0.3124	
Model test	F(34, 3,426) = 16.66; p = 0.0000		F(34, 77,198) = 1,032.81; p = 0.0000	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
\# Significant at the 10% level

Table 6.6. Results of selection-bias corrected wage equation estimated in the independent probit selection model with censoring

6.5.3. Trinomial logit selection model

The results of the trinomial logit model are presented in Table 6.7. The co-efficients are interpreted as the natural logarithm of the odds ratio. These relate the odds of choosing to be employed as a nurse or in some other occupation relative to the reference choice in those for whom the variable is present with those for whom it is absent (in the case of the binary variables) or for a one-unit increase in the value of the variable (in the case of the continuous variables). A statistically insignificant co-efficient or a co-efficient with a value equal to zero indicates an odds ratio equal to one. This means that there is no observed association between the variable and the likelihood of choosing to be employed as a nurse or in some alternative occupation relative to the reference choice. A negative co-efficient indicates an odds ratio of less than one which implies a negative association. A positive co-efficient indicates an odds ratio of greater than one, implying a positive association.

Examining first the results for choosing to be employed as a nurse (D = 2) relative to the reference choice (D = 0), we can see that there is a positive relationship between having a nursing qualification, cohabiting and obtaining a first degree and the likelihood of being employed as a nurse. Individuals with health problems that affect paid work, those who are married, those with a pension, those with no educational qualifications and those who live in the South East of England are less likely to be employed as nurses. In terms of the continuous variables a one-unit increase in non-labour income reduces the likelihood of being employed as a nurse, and there are n-shaped relationships between age and years of schooling and the likelihood of being employed as a nurse.

Individuals who are cohabiting or married, who have a postgraduate degree or a first degree or who live in the South East of England are more likely to choose to be employed in an occupation other than nursing (D = 1) relative to the reference choice (D = 0).

	Employed as a nurse (D = 2)		All other workers (D = 1)	
	ψ_2 [1]	Std.Err.	ψ_1 [1]	Std.Err.
Constant	-8.5936*	0.8155	-3.9246*	0.1689
NURSEQUA	5.9305*	0.0787	-0.2499*	0.0430
Age variables				
AGE	0.1499*	0.0206	0.1159*	0.0042
AGE2	-0.0020*	0.0002	-0.0013*	0.0001
Personal characteristic variables				
DISABLE	-2.0736*	0.0882	-1.7600*	0.0188
ETHNIC	-0.1822	0.1372	-0.7669*	0.0328
NONBRIT	0.0600	0.1276	-0.6175*	0.0340
ETHNBRIT	-0.0321	0.2739	-0.0082	0.0665
Family variables				
PCHILD	0.0918	0.0851	-0.1521*	0.0233
COHABIT	0.3987*	0.0990	0.4138*	0.0247
MARRIED	-0.2052*	0.0658	0.1041*	0.0180
Property income variables				
PENSION	-1.4514*	0.1982	-1.1268*	0.0506
NONLABY	-0.0001*	0.00003	-0.0001*	0.000007
Years of education variables				
YED	0.2824*	0.0942	0.3621*	0.0196
YED2	-0.0111*	0.0031	-0.0125*	0.0006
Educational attainment variables				
PGDEG	-0.2655	0.1644	1.0548*	0.0518
DEG	0.3855*	0.0924	0.6239*	0.0312
ALEVEL	0.0860	0.1940	-0.1085*	0.0263
NOQUAL	-2.2137*	0.2974	-0.9732*	0.0179
Regional variables				
SEAST	-0.1369*	0.0568	0.0395*	0.0149
Log likelihood function		-74,819.60		
Restricted log likelihood		-96,356.90		
Model test		χ^2 = 43,074.62; df = 66; sig. = 0.0000		
N		125,778		

[1] The reference group is non-participators (D = 0)
* Significant at the 5% level
[#] Significant at the 10% level

Table 6.7. Results of trinomial logit selection model

Factors negatively associated with the likelihood of choosing D = 1 are having health problems that affect paid work, being non-white and British or white and non-British, having children, receiving a pension, increasing non-labour income, having A levels as the highest qualification or having no qualifications at all. As before there are n-shaped relationships between age and years of schooling and the likelihood of being employed as a nurse.

The selection-bias corrected estimates of the wage equations of nurses and other workers are presented in Table 6.8. The co-efficients are of the expected order of magnitude and are consistent with the previous models. In terms of the years of education variables for nurses and other workers the maximum earnings occur after 16 and 26 years of education, respectively. In terms of the work experience variables earnings are maximised at 26 and 29 years of work experience for nurses and all other workers, respectively. For the educational attainment variables the returns to obtaining a nursing qualification are greater for nurses than other workers (26% versus 19%), though for non-nurses the premia are all higher for PGDEG, DEG and ALEVEL.

In terms of the personal characteristic variables the co-efficients on DISABLE and ETHNIC are positive for non-nurses though they are not statistically significantly different from zero for nurses. White non-British nurses earn on average 10% higher wages than their colleagues and for other workers the effect is slightly greater (13%). Non-white non-British non-nurses earn on average 7% lower wages than other non-nurses.

As in the previous analyses regional effects are less pronounced for nurses than all other workers. Nurses residing in the South East of England earn on average 7% higher wages than other nurses, while for other workers the average effect on lnW is 17%.

In terms of the job characteristic variables, for nurses there is on average a negative relationship between hours worked and wages, and also between NWORKERS and TEMP and wages. For HOURSPW and NWORKERS the opposite effect arises for all other workers on average. There is a wage premium to being employed in a managerial position for both nurses and all other workers (+5% and +15%, respectively).

The selection bias correction terms are statistically significant and negative in both instances.

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	1.1972*	0.2578	1.0356*	0.0772
Years of education variables				
YED	0.1020*	0.0252	0.0682*	0.0069
YED2	-0.0032*	0.0008	-0.0013*	0.0002
Educational attainment variables				
NURSEQUA	0.2298*	0.0247	0.1750*	0.0117
PGDEG	0.1858*	0.0505	0.3029*	0.0179
DEG	0.0814*	0.0250	0.2430*	0.0104
ALEVEL	-0.0617	0.0681	0.0601*	0.0157
NOQUAL	-0.1667	0.1063	-0.0514*	0.0199
Work experience variables				
EXP	0.0206*	0.0024	0.0290*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	0.0712	0.0569	0.1582*	0.0151
ETHNIC	0.0433	0.0378	0.0867*	0.0133
NONBRIT	0.0933*	0.0328	0.1127*	0.0128
ETHNBRIT	0.0110	0.0647	-0.0798*	0.0242
Regional variables				
SEAST	0.0650*	0.0145	0.1591*	0.0056
Job characteristic variables				
HOURSPW	-0.0038*	0.0006	0.0019*	0.0001
MANAGE	0.0522*	0.0146	0.1417*	0.0038
NWORKERS	-0.0363*	0.0163	0.1339*	0.0033
TEMP	-0.0759*	0.0271	0.0092	0.0061
Selection bias variables				
λ	-0.3384*	0.1384	-0.6654*	0.0391
Adjusted R^2	0.1323		0.3016	
Model test	$F(33, 3{,}427) = 16.98$; $p = 0.0000$		$F(33, 77{,}199) = 1{,}033.59$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE

* Significant at the 5% level

[#] Significant at the 10% level

Table 6.8. Results of selection-bias corrected wage equation estimated in the trinomial logit selection model

6.5.4. Quadrinomial logit selection model

The results of the quadrinomial logit model are presented in Table 6.9. In terms of the decision to be employed as a nurse there is a positive relationship between having children and cohabiting and the likelihood of being employed as a nurse relative to the reference choice D = 0. The negative co-efficients on DISABLE, ETHNIC, NONBRIT, MARRIED, PENSION, NONLABY, PGDEG, ALEVEL, NOQUAL and SEAST indicate that all these factors have a negative impact on choosing option D = 3 relative to D = 0. For all other workers (D = 2) the results are similar except for PCHILD (which has a negative impact), MARRIED, PGDEG and SEAST (which now have a positive impact) and DEG (which is statistically significant and exerts a positive effect).

Because of the way in which the sub-samples were defined in terms of the possession of a nursing qualification the decision to choose D = 1 relative to D = 0 may be thought of as examining the factors that influence the likelihood of having a nursing qualification among the non-participating group. Factors that influence negatively the likelihood of being a non-participating nurse are having health problems that affect paid work, having children, being married and receiving a pension. There is a negative relationship between being British and non-white, white and non-British, cohabiting, and having a first or higher degree and the likelihood of choosing D = 1 relative to D = 0.

For all three groups there are non-linear relationships between both age and years of schooling and the likelihood of being in the selected group relative to the reference choice.

	Participating nurses (D = 3)		Participating non-nurses (D = 2)		Non-participating nurses (D = 1)	
	ψ_3 [1]	Std.Err.	ψ_2 [1]	Std.Err.	ψ_1 [1]	Std.Err.
Constant	-25.9419*	0.8338	-4.0275*	0.1709	-16.8345*	1.1886
NURSEQUA [2]						
Age variables						
AGE	0.3521*	0.0151	0.1170*	0.0043	0.2093*	0.0280
AGE2	-0.0039*	0.0002	-0.0013*	0.0001	-0.0017*	0.0003
Personal characteristic variables						
DISABLE	-1.7860*	0.0739	-1.7547*	0.0189	0.2604*	0.0686
ETHNIC	-0.5703*	0.0995	-0.7752*	0.0330	-0.4307*	0.1673
NONBRIT	-0.2438*	0.0893	-0.6271*	0.0341	-0.3620*	0.1670
ETHNBRIT	-0.1185	0.1896	-0.0057	0.0668	-0.0166	0.3312
Family variables						
PCHILD	0.1588*	0.0614	-0.1517*	0.0235	0.3522*	0.1240
COHABIT	0.2683*	0.0695	0.4094*	0.0248	-0.5280*	0.2056
MARRIED	-0.0924#	0.0486	0.1120*	0.0181	0.2582*	0.0811
Property income variables						
PENSION	-1.0898*	0.1743	-1.0850*	0.0513	0.5240*	0.1126
NONLABY	-0.0001*	0.00002	-0.0001*	0.000007	0.000008	0.00002
Years of education variables						
YED	2.2810*	0.1093	0.3707*	0.0199	1.0237*	0.1450
YED2	-0.0732*	0.0037	-0.0126*	0.0006	-0.0298*	0.0049
Educational attainment variables						
PGDEG	-0.4170*	0.1454	1.0145*	0.0521	-1.1219*	0.3044
DEG	-0.0994	0.0755	0.6082*	0.0314	-0.4126*	0.1418
ALEVEL	-2.7050*	0.1827	-0.1383*	0.0263	-34.4990	2927653
NOQUAL	-4.5638*	0.2905	-1.0069*	0.0179	-35.3505	1927307
Regional variables						
SEAST	-0.2606*	0.0422	0.0413*	0.0150	0.0352	0.0715
Log likelihood function			-87,120.56			
Restricted log likelihood			-101,342.3			
Model test			$\chi^2 = 28{,}443.49$; df = 96; sig. = 0.0000			
N			125,778			

[1] The reference group is non-participating non-nurses (D = 0)
[2] NURSEQUA predicts D = 0 and D = 1 perfectly and so is omitted
* Significant at the 5% level
\# Significant at the 10% level

Table 6.9. Results of quadrinomial logit selection model

The results of the wage equations are presented in Table 6.10. In terms of the years of education variables for nurses and other workers the maximum earnings occur after 16 and 19 years of education, respectively. In this model earnings are maximised at 26 years of work experience for both nurses and other workers. In terms of educational attainment the transformed co-efficients suggest that the premium to obtaining a nursing qualification is higher for nurses than other workers (25% versus 14%). The wage premium to obtaining a postgraduate degree is 27% for nurses and 52% for other workers. The premium to obtaining a first degree is also lower for nurses (26% versus 35%). For all workers being disabled has on average a negative impact on wages (-5% for nurses, -9% for other workers), while for being white and non-British the effect is positive (+7%). Being non-white or both non-white and non-British are not statistically significant for nurses, while for other worker these variables are statistically significant (they exert a positive and negative effect, respectively on mean lnW). Nurses living in the South East of England earn on average 7% higher wages than nurses living outside of this area. For other workers the premium is even higher at 17%.

In terms of the job characteristic variables, there are a number of differences between nurses and all other workers. First, nurses' hourly wages are inversely related to the number of hours worked but there is an opposite effect, on average, for all other workers. There is a wage premium to being employed in a managerial position for both nurses and all other workers (6% for nurses, 15% for other workers). For nurses, being employed in a relatively large workplace or being employed on a temporary contract are negatively related to earnings, while for all other workers on average these effects are positive and statistically insignificant, respectively. As in the previous models, in terms of the time trend variables (not shown – see Appendix 6.10), wage increases for nurses over time were not statistically significant after controlling for measurable individual productive characteristics until the late 1990s. There is generally a quarter-on-quarter increase in real wages, on average, for all other workers however.

For nurses the co-efficient on the selection bias correction term is not statistically significant.

This is interpreted to mean that selection bias is not significant for this group.

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.8921*	0.2129	0.4221*	0.0383
Years of education variables				
YED	0.1127*	0.0247	0.1275*	0.0049
YED2	-0.0036*	0.0008	-0.0034*	0.0002
Educational attainment variables				
NURSEQUA	0.2200*	0.0245	0.1351*	0.0100
PGDEG	0.2368*	0.0475	0.4188*	0.0089
DEG	0.1113*	0.0224	0.2969*	0.0062
ALEVEL	0.1510	0.1894	0.1360*	0.0310
NOQUAL	-0.0010	0.2075	0.9725	0.3014
Work experience variables				
EXP	0.0208*	0.0024	0.0264*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.00002
Personal characteristic variables				
DISABLE	-0.0571*	0.0256	-0.0992*	0.0061
ETHNIC	0.0052	0.0338	0.0234*	0.0093
NONBRIT	0.0636*	0.0296	0.0684*	0.0091
ETHNBRIT	-0.0053	0.0643	-0.1389*	0.0203
Regional variables				
SEAST	0.0668*	0.0143	0.1569*	0.0034
Job characteristic variables				
HOURSPW	-0.0037*	0.0006	0.0023*	0.0001
MANAGE	0.0525*	0.0147	0.1402*	0.0038
NWORKERS	-0.0361*	0.0165	0.1367*	0.0033
TEMP	-0.0785*	0.0273	-0.0011	0.0060
Selection bias variables				
λ	-0.0352	0.0286	-0.1786*	0.0049
Adjusted R^2	0.1312		0.3128	
Model test	$F(33, 3{,}427) = 16.83$; $p = 0.0000$		$F(33, 77{,}199) = 1{,}066.31$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
[#] Significant at the 10% level

Table 6.10. Results of selection-bias corrected wage equation estimated in the quadrinomial logit selection model

6.5.5 Tests for the independence of irrelevant alternatives

The results of Hausman and Small-Hsiao tests of the IIA assumption in the trinomial and quadrinomial logit models are presented in Table 6.11. In terms of the Hausman test in most instances the test statistic is negative and therefore fails to meet the assumptions of the test (i.e. that the test statistic is distributed as Chi-square, which is positive). The results are therefore ambiguous.

	Trinomial logit model			
Hausman test				
Omitted choice	H_{IIA} [1]	df	$P > H_{IIA}$	Hypothesis test [2]
Not participate	-16.884	32	-ve [3]	Ambiguous [3]
Participate + nurse	-21.431	32	-ve [3]	Ambiguous [3]
Participate + non-nurse	9.617	32	1.000	Fail to reject H_0
Small-Hsiao test				
Cmitted choice	SH [1]	df	P > SH	Hypothesis test [2]
Not participate	47.987	34	0.056	Reject H_0
Participate + nurse	33.471	34	0.493	Fail to reject H_0
Participate + non-nurse	56.110	34	0.010	Reject H_0
	Quadrinomial logit model			
Hausman test				
Omitted choice	H_{IIA} [1]	df	$P > H_{IIA}$	Hypothesis test [2]
Not participate + non-nurse	-0.0004	2	-ve [3]	Ambiguous [3]
Not participate + nurse	-31.34	60	-ve [3]	Ambiguous [3]
Participate + nurse	-0.432	58	-ve [3]	Ambiguous [3]
Participate + non-nurse	-0.0009	2	-ve [3]	Ambiguous [3]
Small-Hsiao test				
Omitted choice	SH [1]	df	P > SH	Hypothesis test [2]
Not participate + non-nurse	78.416	33	0.000	Reject H_0
Not participate + nurse	76.053	33	0.000	Reject H_0
Participate + nurse	65.729	33	0.001	Reject H_0
Participate + non-nurse	77.250	33	0.000	Reject H_0

[1] Distributed as Chi-square

[2] H_0: the odds of choice j versus choice k are independent of the other alternatives. The significance level is 10%.

[3] The Chi-square test statistic is negative and therefore does not meet the assumptions of the test. The results of the test are therefore ambiguous.

Table 6.11. Results of Hausman and Small-Hsiao tests of IIA assumption

In terms of the Small-Hsiao test the results indicate that in the vast majority of cases H_0 is rejected. This is the case for both models. What this basically means is that the co-efficients in the model do change when certain choices are omitted. The implication is that the IIA assumption is violated in the trinomial and quadrinomial logit models and the independence assumption of the multinomial logit model is not met. In terms of model selection the correct interpretation is therefore that on statistical grounds (rather than in terms of their ability to model individuals' labour market decisions) there is a clear case for preferring the bivariate probit selection models over and above the multinomial logit selection models.

6.6. Results of the decomposition analysis

We turn now to the results of the decomposition analysis, presented in Table 6.12. The observed difference in mean lnW between nurses and all other workers in the data is 0.3589 (nurses receive on average 33% higher hourly wages than other workers). We decompose this observed pay differential into three main components: due to differences in endowments; due to differences in the returns to endowments; and, due to differences in selection bias. The premium to being a nurse is analysed using the characteristics of nurses and also using the characteristics of all other workers. In Appendix 6.14 we also present the results using OLS estimates without correction for selection bias, participation selection bias corrected estimates, occupation selection bias corrected estimates, and double selectivity corrected estimates estimated using an independent probit model. A more detailed decomposition of the results on a variable-by-variable basis is presented in Appendix 6.15. This shows the contribution of each variable included in the regressions to the overall differences in variables and the overall premium.

In terms of the bivariate probit selection model decomposition the differences in endowments is positive (i.e. greater for nurses) and slightly less than the observed difference in mean lnW. The remainder of the difference is explained by a relatively small positive return to endowments (the premium) to being employed as a nurse. What these figures imply is that nurses earn higher wages than other workers but that this difference is explained primarily but not exclusively by their superior labour market and personal characteristics. For example, as discussed above (Table 6.2) nurses have more years of education than other workers, a greater proportion of nurses have a nursing qualification, and a greater proportion of nurses than all other workers possess at least some form of educational qualification. Nurses also on average tend to have more years of experience than other workers, and a larger proportion of nurses play some kind supervisory role in their job. The differences due to selection bias are negligible.

Decomposition using the parameter estimates generated by the independent probit selection model with censoring yields a slightly different interpretation to explaining wage differentials between nurses and other workers. In this instance the differences in endowments is positive if the premium to being a nurse is analysed using the characteristics of nurses and negative if it is analysed using the characteristics of non-nurses. Inspection of Table A6.15.2 in Appendix 6.15 shows that the largest differences in the contribution of individual variables to the overall differences in variables depending on how the premium to being a nurse is analysed is explained by the variables NURSEQUA, MANAGE and NWORKERS. This difference does exist in the previous model (see Table A6.14.1) but the differences are compensated by the effects of the other variables so that the differences in variables is positive in both cases. Just as in the uncensored model the premium to being a nurse is positive. The selection bias effects in this model are also small (0.0508).

A different set of results is obtained from the decomposition of wage differentials using the multinomial logit selection models. It is important to bear in mind however that these models do violate the IIA assumption of the multinomial logit model and should therefore be treated with caution. In the case of the trinomial logit selection model the differences in endowments is positive and greater than the observed differences in returns to endowments (which are negative). The remaining difference in mean lnW is explained by the differences due to selection bias. A similar set of results is generated using the quadrinomial logit selection model. In this case the differences in variables is also positive and offset by the negative return to endowments. The differences due to selection bias again contribute importantly to the observed differences in mean lnW.

	Premium to being a nurse analysed using characteristics of nurses	Premium to being a nurse analysed using characteristics of non-nurses
Bivariate probit selection model		
Differences in variables (= differences in endowments)	0.2597	0.2887
Premium (= differences in returns to endowments)	0.0986	0.0696
Differences due to occupation and participation selection bias	0.0006	0.0006
Observed difference in mean lnW	0.3589	0.3589
Independent probit selection model with censoring		
Differences in variables (= differences in endowments)	0.1668	-0.0432
Premium (= differences in returns to endowments)	0.1413	0.3513
Differences due to occupation and participation selection bias	0.0508	0.0508
Observed difference in mean lnW	0.3589	0.3589
Trinomial logit selection model		
Differences in variables (= differences in endowments)	0.2968	0.2715
Premium (= differences in returns to endowments)	-0.2529	-0.2276
Differences due to selection bias	0.3150	0.3150
Observed difference in mean lnW	0.3589	0.3589
Quadrinomial logit selection model		
Differences in variables (= differences in endowments)	0.0589	0.2270
Premium (= differences in returns to endowments)	-0.2618	-0.4299
Differences due to selection bias	0.5618	0.5618
Observed difference in mean lnW	0.3589	0.3589

Table 6.12. Results of the decomposition analysis

6.7. Conclusion

In Chapter 6 we have expanded on the 'single-selectivity' models of Chapter 5 and analysed using four models (bivariate probit selection model, independent probit selection model with censoring, trinomial logit selection model and quadrinomial logit selection model) the determinants of wages for nurses and other workers in Great Britain using QLFS data from Spring 1997 to Autumn 2000, correcting for potential selection bias. Because they correct simultaneously for two forms of selection bias these models are referred to as 'double selectivity models'. This is a novel approach: in the review section of the chapter we find that these types of model are extremely rare in the literature, and there has been only a single application to the (US) nursing labour market (Botelho et al., 1998), though the model was used in a different context.

The four statistical models we estimate model in different ways the participation and the occupation selection decisions faced by the individual. For the two models based on the bivariate probit the assumption is that individuals' make two decisions between two choices (to participate or not and to work as a nurse or not). We include in the wage equations for nurses and other workers two selection bias correction terms that capture the effects of both the participation and the occupation selection decisions. Two models of this type are estimated: one with and one without what Greene (2000) calls censoring, where the observed variables in the bivariate probit model are censored in some way. The difference is that in the uncensored model occupation selection decisions are not conditional on the decision to participate. This allows for the possibility that individuals who do not work may still consider themselves to be 'in' an occupation group in some sense (there may be non-participating nurses, for example). In the censored model the occupation selection decision is observed only when the individual decides to participate.

The second type of model we employ is the multinomial logit selection model. This also entails including in the wage equations for nurses and other workers selection bias correction terms that capture the effects of both the participation and the occupation selection decisions. The main difference is that in the multinomial logit model there is a single decision between more than two alternatives. We estimate two models of this type. In the first model – a three-option model (called the trinomial logit selection model) – there are three alternatives: to participate in the labour market as a nurse; to participate in the labour market in an occupation other than nursing; or, to not participate in the labour market at all. We also estimate a four-option model (the quadrinomial logit selection model) for which the alternatives are to participate in the labour market as a nurse, to participate in the labour market in an occupation other than nursing, to be a nurse and to not participate in the labour market, or to be a non-nurse and to not participate in the labour market. The distinction is similar to that in the bivariate probit models where in the first case the occupation selection decision is conditional upon the decision to participate and in the other it is not.

We discuss at length in the text the characteristics and relative merits of each model. The outcome is that each model involves a different view of how individuals perceive and make their participation and occupation selection decisions. The important point is that a priori all four models are equally valid: from an economic point of view in terms of their relevance to individuals' labour market decisions it is difficult to conclude that one model is better than any other. However, on statistical grounds we find that because the multinomial logit models violate the independence of irrelevant alternatives (IIA) assumption the results of the bivariate probit models are superior and should be preferred.

There are four important and useful outcomes of the analysis. First, we identified the factors in the wage equation that affect nurses' earnings. As predicted by the Mincerian model there is a concave relationship between nurses' earnings and their years of full-time education and their earnings and years of work experience. Important factors that in general positively affecting nurses' earnings are: possessing a nursing qualification; obtaining a first degree or postgraduate degree; being of non-British nationality; living in the South East of England; and, working as a supervisor, manager or foreman. Factors that in general have a negative influence of nurses' hourly wages are: working longer hours; working at a workplace with 25 or more staff; and, having a non-permanent or temporary job. There is generally much agreement across the different models used in this chapter and in relation to the single selectivity models estimated in Chapter 5 in terms of the statistical significance, sign and order of magnitude of the co-efficients. This emphasises the plausibility and robustness of the findings.

The second useful finding concerns the importance of assuming independent disturbances in the multinomial logit model. From the point of view of statistical estimation it is useful that the odds ratios estimated by multinomial logit do not depend on the other choices. However, from a behavioural point of view the IIA assumption is not very attractive. We test for the validity of the assumption in both the trinomial and quadrinomial logit models using a Hausman test (Hausman and McFadden, 1984) and the Small-Hsiao test (Small and Hsiao, 1985). We find that the IIA assumption is violated in both models. In terms of model selection the correct interpretation is therefore that on statistical grounds there is a clear case for preferring the results of the bivariate probit selection models over and above those of the multinomial logit selection models. This result confirms the importance of testing the IIA assumption when interpreting the result of multinomial logit models. [25]

The third important finding concerns the estimation procedure for the bivariate probit selection models. The correct a priori starting point is that the participation and occupation selection decisions may be dependent (made simultaneously). The two equations are estimated by bivariate probit and the correlation co-efficient ρ_v measuring the correlation between the disturbances in the two equations is examined. If ρ_v is statistically significantly different from zero the correct procedure is to estimate the two equations simultaneously in a bivariate probit model. If on the other hand ρ_v is not statistically significant it is appropriate to estimate the two selection equations separately as independent probit models. On this basis for the uncensored model the bivariate probit framework is correct. For the censored model however the independent probit model is more appropriate. If the participation and occupation selection equations in the uncensored model are instead estimated (incorrectly) as two independent probits the relative importance of the selection bias variables is altered. For example, in the rejected independent probit model (see Appendix 6.9 Table A.6.9.2) both the selection bias variables for nurses are statistically significantly different from zero, which is not the case in the bivariate probit model. Additionally a different decomposition is obtained (the differences due to participation and occupation selection bias is negative – see Appendix 6.12). This raises doubts concerning the results of previous double selectivity studies (very small in number) where the independence of the two selection decisions is simply assumed and not tested (see for example Shields and Wheatley Price, 1998, and Botelho et al., 1998).

[25] As revealed from the literature review conducted earlier in this chapter a single study to date (Botelho et al., 1998) has used a multinomial framework to analyse the earnings of nurses, though in a slightly different context. A multinomial logit selection model was used to analyse the effect of the method of entrance into the nursing

The fourth important and useful finding concerns the results of the decomposition analysis. Nurses in the data receive on average 33% higher hourly wages than other workers. This difference is attributable in part to differences in nurses' labour market and personal characteristics, in particular their greater number of years of education and superior educational attainment, their greater propensity to be employed in a supervisory role and their greater number of years of work experience. In addition to generally having superior labour market endowments, in the bivariate probit selection model and the independent probit selection model with censoring the returns to these endowments are also on average higher for nurses than other workers. In the trinomial logit selection model and the quadrinomial logit selection model the returns to endowments are on average lower. Because the multinomial logit models violate the IIA assumption the results should be treated with scepticism and the results generated by the bivariate probit selection models should be preferred. The upshot is that the wage premium to being employed as a nurse is positive. This means that the average nurse would earn lower wages if paid according to the pay structure of other workers. After controlling for differences in individual and productive characteristics and selection bias nurses are paid higher wages than other workers.

Note that this finding is in direct contrast with the findings from the single selectivity models presented in Chapter 5. It is important to remember that the statistical models utilised in Chapters 5 and 6 have been developed from an economic model earnings. The basic model is the Mincerian earnings function. Underpinning the selection bias problem addressed in the statistical models are the economic theories of individual labour supply (relevant to participation selection bias) and occupation selection (occupation selection bias), discussed in

---

profession in the US (associate degree, diploma in nursing, or bachelor of science in nursing) on earnings. No

Chapter 4. Only in the double selectivity models are these economic models fully taken into account. Thus from an economic perspective the double selectivity models are to be preferred. This raises concerns for the vast majority of studies conducted in this area (see the literature review in Chapter 4), which usually concentrate on the more straightforward 'single selectivity' framework correcting mainly for potential participation selection bias only. The findings of this thesis suggest that this approach may be flawed because *from an economic theory point of view* the problem of occupation selection bias is ignored and because *from a statistical point of view* correcting also for occupation bias in the same model may modify the overall results. It was informative to conduct the single selectivity models in Chapter 5 in order to show that discrepancies may occur between the two types of model.

Another interesting finding arising from a comparison of the decomposition analyses in Chapters 5 and 6 is that the results of the basic OLS model in Chapter 5 (which makes no adjustment for selection bias at all) are the same sign as the bivariate probit selection model results in this chapter. A simple comparison of Table 5.13 in Chapter 5 and Table 6.12 in Chapter 6 reveals that there is a positive premium to being employed as a nurse using the simple OLS estimates and the bivariate probit double selectivity models. This raises a question as to the value added from using the rather more sophisticated statistical models of Chapter 6. Put simply, if the complex double selectivity models yield the same outcome as the simple OLS model vis-à-vis the premium to being employed as a nurse then why utilise the more complex model in the first place? The answer is that a priori it was uncertain as to what the actual outcome of the double selectivity models would be. There was no reason to expect the results to necessarily be the same. More importantly, as discussed in the previous paragraph according to the two economic models (of individual labour supply and occupation

---

test was conducted for the independence of irrelevant alternatives.

selection) which provide the economic framework driving the statistical selection bias models one would expect a priori there to be potential selection bias effects arising from an individuals' participation and occupation selection decisions. In other words, prior to the commencement of the analysis an adjustment for both forms of potential selection bias was justified based on the economic theory.

In summary, the double selectivity models are to be preferred over the more usual single selectivity models and the simple OLS model because they control for both forms of selection bias which, based on economic theory, may be problematic in earnings function analyses. The upshot is that on the basis of the economic theory underpinning labour market behaviour a double selectivity approach is advocated.

The important empirical finding from this chapter is that contrary to the findings in Chapter 5, when *both* participation *and* occupation selection bias are controlled for there is evidence of a premium to being employed as a nurse in Great Britain. We therefore conclude from the analysis that after controlling for differences in labour market and individual productive characteristics there are financial benefits to being employed as a nurse in Great Britain relative to other occupations.

CHAPTER 7

CONCLUSIONS OF THE ANALYSIS AND SOME POLICY IMPLICATIONS

7.1 Introduction

In this thesis we have examined in much detail the earnings of nurses in Great Britain. There were two general aims: to delineate the factors that affect nurses' earnings; and, to examine the nature and magnitude of wage differentials between nurses and other workers. We noted at the outset that a detailed analysis of nurses' earnings is important for a number of reasons: the size of the nursing workforce (the NHS currently employs some 415,000 whole-time equivalent nurses); the size of the nursing paybill (£10 billion in 2000 representing 20% of NHS expenditure); current recruitment and retention problems (the current vacancy rate is around 2%); and, from an academic point of view, due to the unusual and complex nature of the market.

To meet the general aims the preceding six chapters of the thesis have provided a comprehensive examination of the earnings of qualified nurses working in the NHS in Great Britain. In the first two chapters the characteristics of the labour market in which qualified nurses work was discussed and the process by which nurses pay is determined was described in considerable detail. The exposition was supplemented by a review of nurses' earnings over the period 1975 to 2000. In the latter four chapters a thorough and wide-ranging analysis of nurses' earnings was conducted, involving estimation of the internal rate of return and net present value to becoming a nurse, and the construction of earnings functions for nurses and other workers with appropriate adjustment for selection bias.

In this the final chapter we summarise the earlier findings and then discuss some policy implications of the analysis. We focus on two aspects: the method of nurses pay determination and the consequent bargaining strategies of the Staff Side and Management Side; and, some tentative suggestions for reducing the current nursing shortage. We then discuss some study limitations and offer some suggestions for future research.

7.2. Main findings

We began our analysis by examining the mechanisms by which nurses' pay is determined and the structure of the nursing labour market. We identified the determinants of the demand and supply of nursing labour and then studied their interaction. On the demand side the government plays a crucial role in setting the NHS budget and, as a consequence, defining the expenditure limits within which wage and employment decisions are constrained. Also important is the Pay Review Body that determines levels of pay, but whose recommendations are influenced heavily by the monopsony power exerted by NHS employers. Through the allocation of funds to the NHS by the government health care providers each year are effectively allocated a budget with which to meet the cost of providing all contracted health services. Thus at the given wage rate set by the Pay Review Body the government effectively determines the number of nurses employed by setting the size of the nursing paybill which acts as a budget constraint on the number of employed workers at the given wage. The implication is that employers base their employment decisions on the interaction of wages and budget. There is no demand curve for nursing labour in the traditional sense. Instead there is a hard budget constraint (an isoexpenditure curve) limiting the maximum number of nurses that may be feasibly employed at any give wage.

On the supply side the decisions of individuals in terms of joining and leaving the profession are paramount in determining the state of the labour market. This defines the stock of nursing labour. The stock will remain constant if the numbers of joiners and leavers is equal, though clearly it may also increase or decrease over time. Recent estimates by the Pay Review Body suggest that the number of qualified nurse joiners is greater than the number of leavers (by approximately 2,000 whole-time equivalents). The majority of joiners are transfers from within the NHS and newly qualified nurses. The most important reasons for leaving are to transfer to other NHS units, to retire or to work in the private sector.

The numbers of joiners and leavers is a function of individuals' labour market participation and occupation selection decisions. The first decision is whether or not to work in the labour market at all, and the second is whether or not to choose to be employed specifically as an NHS nurse. Individuals will choose to participate if the offered wage is greater than their reservation wage. Utility-maximising individuals base their occupation selection decisions on the relative expected financial and non-financial costs and benefits of alternative occupations. If the expected benefits are greater than the expected costs the individual will choose to join (in the case of potential new entrants) or will remain working in (in the case of current workers) a particular occupation. If the expected costs are greater than the expected benefits then the individual will make an alternative choice.

We then discussed the process by which nurses' pay is determined. The Pay Review Body since its formation in 1983 has been responsible for setting salary levels for nurses in Great Britain. It reviews evidence from the Staff Side, the Management Side and from the wider economy in terms of labour market conditions. Taking into account a variety of issues in its deliberations such as recruitment and retention and fairness and comparability the Pay

Review Body makes its own independent recommendations on nurses' pay that the government is obliged to accept unless "there are clear and compelling reasons for not doing so". Nurses have realised salary increases each year under the Pay Review Body system. In some years these have been substantial, though the effects have been limited somewhat by government intervention delaying or staging full implementation. While the Pay Review Body takes a number of issues into account in its deliberations (affordability of potential pay rises, recruitment and retention, fairness and comparability, morale and motivation, and productivity and workload) we find evidence to suggest that in the past the issue of affordability stressed frequently by the monoponistic employers is given much prominence, though more recently it is recognised that recruitment and retention are of prime importance. The outcome is that at least up until recently the market wage rate has been set by the Pay Review Body more in line with the preferences of the Management Side, as opposed to the higher wage levels preferred by the Staff Side.

In Chapter 2 we conduct a review of nurses' earnings. Using data derived from the New Earnings Survey we examine trends in nurses' mean earnings over time for the period 1975 to 2000. We find they have generally increased year on year. This might occur for two broad reasons: due to increases in nurses' salary scales, which arise as a result of the recommendations made by the Pay Review Body; and, due to experience as nurses move to a higher point on the pay scale. This second reason highlights the importance of labour market experience on earnings – found to be an important explanatory variable in the wage equations estimated in Chapters 5 and 6.

An important finding in the review which is borne out by the analyses of Chapters 3 to 6 is that female nurses receive higher mean earnings than the comparator groups (female non-manual workers and female public sector non-manual workers). The upshot is that the vast majority of nurses earn on average higher wages than workers in comparable occupations. We also find that not only do nurses receive higher mean earnings but also that they have enjoyed some of the largest increases in real earnings over the period 1975 to 2000.

Taken in conjunction the main findings of the first two chapters are summarised as follows. First, nurses' pay is determined by a complex mechanism involving interaction between nurses, employers and the Pay Review Body. While the Pay Review Body considers a number of issues in its deliberations in recent years evidence suggests that the issue of affordability has been an important principle on which it has based its recommendations. Increasingly the issues of recruitment and retention are becoming prominent. Second, evidence suggests that the outcome of this process is that nurses' wages are set at or below the constrained equilibrium, more in line with the preferences of the monopsonistic employers. Third, as a consequence the labour supply decisions of nurses and potential new entrants into the profession in terms of their participation and occupation selection decisions dominate. Fourth, there is a shortage of qualified nurses, though this shortage arises for a variety of reasons. Fifth, even though wages are set below the constrained equilibrium nurses on average receive higher earnings than workers in comparable occupations. These findings set the scene for the analyses of Chapters 3 to 6.

In Chapter 3 we examine the lifetime costs and benefits of being employed as a nurse in Great Britain for the period 1991-1996 and measure the private net present value and the private internal rate of return. This goes beyond the review of earnings in Chapter 2 where we focus on mean earnings of nurses and other workers at a specific point in time. In Chapter 3 we provide a more comprehensive measure of the returns to being employed as a nurse and look instead at mean lifetime earnings in nursing relative to opportunity cost occupations (the base case is non-manual workers). From the literature review we find that while the number of studies measuring the attractiveness of investments in human capital in this way is massive there has to date been no comparable study of the returns to nursing in Great Britain.

The IRR and NPV calculations are made using the standard equations inputted with data from the New Earnings Survey and the British Household Panel Survey. Basic age-earnings profiles are adjusted for mortality, unemployment, other causes of economic inactivity, and discontinuation from training. In terms of the private internal rate of return we find that this is high for nurses in Great Britain relative to other occupations. We also show however that using the internal rate of return criterion is inappropriate when a comparison of mutually exclusive investments (occupations) is required (e.g. becoming a nurse, becoming a teacher or obtaining a degree) and there exists a crossover marginal time preference rate. This is in fact shown to be the case here and therefore the net present value criterion is preferred. On this basis we find that nursing is the preferred option on financial grounds for individuals with an MTPR of 8%-12% or more. The implication of this finding, which is consistent with the review of earnings in Chapter 2, is that on financial grounds in terms of their relative earnings there is a rationale for choosing to be employed as a nurse in Great Britain.

From Chapters 2 and 3 we show that there are financial returns to being employed as a nurse. In Chapters 4 to 6 we develop our analysis of nurses' earnings further and look at two specific issues. We examine the individual characteristics that affect nurses' earnings and then we define the nature and magnitude of the wage differential between nurses and other workers. This goes beyond the analysis of the previous chapters because in this instance we examine first why nurses' earnings are of the magnitude they are and then we examine the causes of the observed earnings differentials.

In Chapter 4 we develop a theoretical model of earnings which is then estimated in Chapter 5. The model is based on the work of Mincer (1974) who shows in a framework suitable for econometric estimation that two important factors driving earnings are the amount of compulsory and non-compulsory education received and years of work experience. The extended earnings functions that we estimate are of the form:

$$\ln W_{ji} = \beta_j X_{ji} + U_{ji} \qquad [7.1]$$

where W is wages, X is a matrix of individual human capital characteristics (including years of schooling and its square and post-school work experience and its square) and other exogenous socio-economic variables affecting wages, β is a vector of unknown parameters to be estimated, U is a normally distributed error term with zero mean and constant variance, σ_U^2. The subscript j allows us to estimate separate equations for nurses (j = n) and other workers (j = o).

We supplement the Mincerian model with an examination of labour market participation and occupation selection decisions. This is relevant because as discussed in Chapter 1 it is the supply-side decisions of individuals in terms of the participation decision and the occupation selection decision that determines the state of play in the nursing labour market. More importantly from an estimation point of view Heckman (1979) has shown that failure to account for the self-selected nature of the decision to participate in the labour market and the decision to choose to be employed in a particular occupation leads to biased estimates of the Mincerian earnings function. In terms of the participation selection bias problem we adjust for the possibility that employees may differ systematically in unobservable characteristics from those who choose not to participate. This would be the case if due to their unobserved ability, motivation or personal circumstances a non-participator's reservation wage were greater than the wage offered by employers. This might arise because the individual would otherwise earn relatively low wages (they have relatively low offered wages) and therefore the sample of observed wages would be biased upwards. Alternatively individuals who choose not to work might have earned higher wages than those who do choose to work but they have an even higher reservation wage – in which case the sample of observed wages is biased downwards. A further possibility is occupation selection bias which might occur for example if individuals self-select into occupations in which they have a comparative advantage in terms of natural ability and motivation. This means that simple comparison of the earnings of nurses and other workers may be a biased estimate of the returns to being employed as a nurse for any given individual. The problem is that comparing the actual mean earnings of nurses and individuals employed in other occupations may overstate or understate the true returns to being employed as a nurse if an individual employed as a nurse would earn higher or lower wages if employed in another occupation than someone already employed in that occupation or vice versa. We therefore augment the analysis by correcting separately for potential participation selection bias and occupation selection bias using the Heckman two-step procedure. This involves estimating by probit participation equations and occupation selection equations for nurses and other workers and then including in the wage equations the selection bias correction terms (the inverse Mills ratios – the λ's) that capture the propensity to participate in the labour market and the propensity to be employed as a nurse. The outcome is that wage equations [7.2] and [7.3] are estimated which correct for potential participation selection bias and occupation selection bias, respectively.

$$\ln W_{ji} = \beta_j X_{ji} + \beta_{\lambda(p)j}\lambda(p)_{ji} + \varepsilon_{ji} \qquad [7.2]$$

$$\ln W_{ji} = \beta_j X_{ji} + \beta_{\lambda(nu)j}\lambda(nu)_{ji} + \varepsilon_{ji} \qquad [7.3]$$

$\lambda(p)$ is the inverse Mills ratio denoting a correction for participation selection bias and $\beta_{\lambda(p)}$ is its co-efficient. $\lambda(nu)$ and $\beta_{\lambda(p)}$ are defined analogously and denote an adjustment for occupation selection bias.

Having outlined the basic model in Chapter 4 we then review the literature to date on earnings function for nurses. Previous studies are based on analyses of nursing labour supply, analyses of monopsony power in the nursing labour market, analyses of the returns to different types of nursing education, and analyses of factors affecting the growth in wage rates of nurses over time. Unfortunately, while the coverage of previous work in this area is substantial, as evidenced by the number of studies, it concentrates primarily on the US nursing labour market. The results do not apply to Great Britain since the method of entry into the profession, the structure of the labour market and the method of pay determination are different. It is also notable that the US studies suffer frequently from selection bias problems of the kind alluded to above and are often mis-specified.

There has to date been a single earnings function analysis for nurses in the NHS in Great Britain (Phillips, 1995) conducted as part of a wider analysis of nursing labour supply. This study is based on 1980 data, however, and so is quite dated. For example, as discussed in Chapter 2, from 1980 to the present the method by which nurses' pay is determined has changed substantially (the Pay Review Body was established in 1983) and the profession is structurally different following the clinical regrading exercise in the late 1980s. Additionally the estimated model does not include a number of potentially important explanatory variables.

Having justified and defined the economic model and the statistical models to be estimated, in Chapter 5 we estimate earnings functions for nurses and other workers using the methods outlined above. We in fact estimate five statistical models that involve the estimation of wage equations for nurses and other workers with corrections for participation selection bias and occupation selection bias using the Heckman two-step procedure. A summary of the estimated models is presented in Table 7.1. The data to which the models are applied are taken from the Quarterly Labour Force Survey, a random survey of representative households in Great Britain. The final sample from is taken from the period 1991 to 2000 and consists of 247,774 females aged 18 to 60 years of whom 8,878 are employed as NHS nurses.

Model	Structure	Sample	Estimation	Wage equation [1]	Selection bias effect
1	Wage equation with dummy variable for whether or not an individual is employed as a nurse	Workers only	OLS	$\ln W_i=\beta X_i+\beta_n Nu_i+U_i$	n/a
2	Participation equation and wage equation with dummy variable for whether or not an individual is employed as a nurse	Workers and non-workers (participation equation), workers only (wage equation)	Heckman two-step procedure	$\ln W_i=\beta X_i+\beta_n Nu_i+\beta_{\lambda(p)}\lambda(p)_i+\varepsilon_i$	$\lambda(p)$ statistically significant and negative
3	Separate wage equations for nurses and all other workers	Workers only	OLS	$\ln W_{ni}=\beta_n X_{ni}+U_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+U_{oi}$	n/a
4	Separate participation equations and wage equations for nurses and all other workers	Workers and non-workers (participation equation), workers only (wage equations)	Heckman two-step procedure	$\ln W_{ni}=\beta_n X_{ni}+\beta_{\lambda(p)n}\lambda(p)_{ni}+\varepsilon_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+\beta_{\lambda(p)o}\lambda(p)_{oi}+\varepsilon_{oi}$	$\lambda(p)_n$ not statistically significant, $\lambda(p)_o$ statistically significant and negative
5	Occupation selection equation, and separate wage equations for nurses and all other workers	Workers only	Heckman two-step procedure	$\ln W_{ni}=\beta_n X_{ni}+\beta_{\lambda(nu)n}\lambda(nu)_{ni}+\varepsilon_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+\beta_{\lambda(nu)o}\lambda(nu)_{oi}+\varepsilon_{oi}$	$\lambda(nu)_n$ statistically significant and positive, $\lambda(nu)_o$ not statistically significant.

[1] X_j = years of full-time education, years of full-time education squared, as a nursing qualification, highest qualification is a postgraduate degree, highest qualification is a first degree, highest qualification is A level, has no qualifications, years of experience with current employer, years of experience with current employer squared, has health problems affect paid work, non-white ethnic group, non-British nationality, non-white and non-British, lives in the South East of England, total usual hours worked per week, employed as a supervisor, manager or foreman, 25 or more workers at workplace, job is non-permanent or temporary, plus 31 time trend dummy variables

Table 7.1. Main features of Models 1-5 in Chapter 5

There are three main outcomes from the analysis. First, we determined the factors in the wage equation that affect nurses' earnings. Factors positively affecting nurses' earnings are: years of full-time education (there is a concave relationship); possessing a nursing qualification; obtaining a first degree or postgraduate degree; years of work experience (concave relationship); being of non-British nationality; living in the South East of England; and, working as a supervisor, manager or foreman. Factors that have a negative influence on nurses' hourly wages are: possessing A levels as the highest qualification; having no qualifications; having health problems that affect paid work; working longer hours; working at a workplace with 25 or more staff; and, having a non-permanent or temporary job. Ethnic group was found to have a negligible influence on nurses' earnings (the co-efficient on this variable was not statistically significant). These results were consistent across the estimated models. We also found from Model 4 that the co-efficient on the participation selection bias correction term $\lambda(p)_n$ was not statistically significant, indicating that for nurses participation selection bias is not significant (see Table 7.1). In terms of potential occupation selection bias (Model 5), the co-efficient on this variable $\lambda(nu)_n$ was found to be statistically significant and positive. [26]

From an estimation point of view the second main finding was the importance of the marginal effects in the wage equation. In the Heckman two-step procedure the full marginal effect on wages of variables that appear as regressors in both the participation/occupation selection equation and the wage equation consists of two components. There is the direct effect on the mean of lnW, which is the co-efficient β in the wage equation. In addition, for independent variables that also appear in the participation/occupation selection equation an

[26] Note that the statistical insignificance of the inverse Mills ratios does not mean that individuals' participation and occupation selection decisions are not important in the nursing labour market, as discussed above, only that there are no selection bias effects.

indirect effect on lnW will also be exerted through their influence on λ. We estimate the full marginal effects. What is important is that the magnitude, sign and statistical significance of the marginal effects are for many variables different from those of the direct effect given by the relevant co-efficient in the wage equation. This calls into question the conclusions of many earlier earnings function studies that utilise the Heckman two-step approach where the issue is often overlooked.

The third main outcome from Chapter 5 pertains to the nature and magnitude of the earnings differential between nurses and other workers. We find that nurses in the sample are paid on average higher wages than workers in all other occupations combined. The mean real hourly wages of nurses and all other workers are £7.36 (Std. Dev. £2.96) and £5.49 (Std. Dev. £3.50), respectively. The difference in mean real hourly wages (£1.87 – nurses receive on average 34% higher wages than all other workers) is statistically significant at conventional levels ($p<0.0001$, 95% confidence interval £1.80 to £1.95). Using the algebraic method developed by Oaxaca (1973) we decompose the observed difference in mean ln wages (0.3648) into differences in labour market endowments and differences in the returns to these endowments. The decomposition is informative for the following reason. We wish to compare the earnings of nurses to the earnings of other workers in order to determine whether there is a financial return to being employed as a nurse. One option is to compare the mean earnings of nurses and the mean earnings of all other workers. 'All other workers' however includes both non-manual workers and manual workers some of whom have entirely different years of education, qualifications, and job characteristics to nurses. Thus aside from the potential selection bias problem a raw comparison is not necessarily informative because we may not be comparing like with like. It would be unsurprising to find that a qualified nurse with 15 years experience earns higher hourly wages than cleaner with one year of experience and no post-compulsory schooling and no qualifications, for example.[27] In Chapter 5 (and 6) the approach adopted is to compare nurses to all other workers in order to utilise the full sample of the available data. The comparison is not problematic because it is possible to disaggregate the observed earnings differential into an endowment component – in which higher earnings are observed due to superior labour market endowments such as schooling, qualifications, etc. – and a premium component which arises due to differences in the returns to endowments. This second effect allows for the possibility that, for example, the impact on earnings of having more experience or having a postgraduate qualification is different for nurses and other workers. In the comparison we wish to control for the differences in endowments. Having then effectively removed the endowment component from the earnings differential we examine the premium. We can conclude that nurses are paid more than other workers if the premium is positive. The implication in this case is that the average nurse would earn lower wages if paid according to the pay structure of other workers. The opposite interpretation is true if the premium is negative. In the selection-bias corrected estimates of Chapter 5 we find that the endowment effect is positive and the premium is negative. The interpretation is that nurses are paid higher wages than other workers but that this difference is due exclusively to differences in nurses' superior labour market and personal characteristics, in particular their greater number of years of education and superior educational attainment, their greater propensity to be employed in a supervisory role and their greater number of years of work experience. The returns to labour market endowments are on average lower for nurses than other workers – the premium is negative. Put another way, after controlling for differences in individual and productive characteristics and selection bias nurses are in fact paid lower wages than other workers.

[27] Note that in Chapters 2 and 3 we compensated for this effect by using non-manual workers as the baseline

This is an important finding and has clear implications for reducing the current nursing shortage. However, it should be treated with caution because while the analysis controls for the effect of potential participation selection bias and occupation selection bias the corrections were made individually in separate models. This leads us to Chapter 6 where we construct extended earnings functions for nurses and other workers in Great Britain correcting jointly for *both* participation selection bias *and* occupation selection bias *in the same model*. Because they correct simultaneously for two forms of selection bias these models are referred to as 'double selectivity models'. This is a novel approach: in the review section we find that these types of model are extremely rare in the literature, and there has been only a single application to the (US) nursing labour market (Botelho et al., 1998), though the model was used in a different context.

We estimate four statistical models using a bivariate probit framework and a multinomial logit framework which treat in different ways the effects of both the participation and the occupation selection decisions. For the two models based on the bivariate probit the assumption is that individuals' make two decisions between two choices (to participate or not and to work as a nurse or not). We include in the wage equations for nurses and other workers two selection bias correction terms that capture the effects of both the participation and the occupation selection decisions. The estimated wage equations are as follows:

$$\ln W_{ji} = \beta_j X_{ji} + \beta_{\lambda(p)j}\lambda(p)_{ji} + \beta_{\lambda(nu)j}\lambda(nu)_{ji} + \varepsilon_{ji} \qquad [7.4]$$

Two models of this type are estimated: one with and one without what Greene (2000) calls censoring, where the observed variables in the bivariate probit model are censored in some

---

comparator/opportunity cost group.

way. The difference is that in the uncensored model occupation selection decisions are not conditional on the decision to participate. This allows for the possibility that individuals who do not work may still consider themselves to be 'in' an occupation group in some sense (there may be non-participating nurses, for example). In the censored model the occupation selection decision is observed only when the individual decides to participate.

The second type of model we employ is the multinomial logit selection model. This also entails including in the wage equations for nurses and other workers selection bias correction terms that capture the effects of both the participation and the occupation selection decisions. The main difference is that in the multinomial logit model there is a single decision between more than two alternatives. We estimate two models of this type. In the first model – a three-option model (called the trinomial logit selection model) – there are three alternatives: to participate in the labour market as a nurse; to participate in the labour market in an occupation other than nursing; or, to not participate in the labour market at all. We also estimate a four-option model (the quadrinomial logit selection model) for which the alternatives are to participate in the labour market as a nurse, to participate in the labour market in an occupation other than nursing, to be a nurse and to not participate in the labour market, or to be a non-nurse and to not participate in the labour market. The distinction is similar to that in the bivariate probit models where in the first case the occupation selection decision is conditional upon the decision to participate and in the other it is not. The generalised wage equation is given by:

$$\ln W_j = \beta_j X_{ji} + \beta_{\lambda j}\lambda_{ji} + \eta_{ji} \qquad [7.5]$$

where λ_j denotes an adjustment for the self-selected nature of choice j based on the odds of choosing choice j relative to the reference choice.

We discuss at length in Chapter 6 the characteristics and relative merits of each model. The outcome is that each model involves a different view of how individuals perceive and make their participation and occupation selection decisions. In terms of their relevance to individuals' labour market decisions a priori it is difficult to conclude that one model is better than any other. However, on statistical grounds we find that because the multinomial logit models violate the independence of irrelevant alternatives (IIA) assumption the results of the bivariate probit models are superior and should be preferred (see below).

A summary of the estimated models is presented in Table 7.2. The data to which the models are applied were taken from the Quarterly Labour Force Survey.

Model	Structure	Sample	Estimation	Wage equation [2]	Selection bias effect
Bivariate probit selection model	Participation equation, uncensored occupation selection equation, and separate wage equations for nurses and all other workers	Workers and non-workers (participation and occupation selection equations), workers only (wage equation)	Bivariate probit (participation and occupation selection equation), OLS (wage equation)	$\ln W_{ni}=\beta_n X_{ni}+\beta_{\lambda(p)n}\lambda(p)_{ni}+\beta_{\lambda(nu)n}\lambda(nu)_{ni}+\varepsilon_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+\beta_{\lambda(p)o}\lambda(p)_{oi}+\beta_{\lambda(nu)o}\lambda(nu)_{oi}+\varepsilon_{oi}$	$\lambda(p)_n$ and $\lambda(nu)_n$ not statistically significant, $\lambda(p)_o$ statistically significant and negative, $\lambda(nu)_o$ not statistically significant
Independent probit selection model with censoring [1]	Participation equation, censored occupation selection equation, and separate wage equations for nurses and all other workers	Workers and non-workers (participation equation), workers only (occupation selection and wage equations)	Independent probits (participation and occupation selection equation), [1] OLS (wage equation)	$\ln W_{ni}=\beta_n X_{ni}+\beta_{\lambda(p)n}\lambda(p)_{ni}+\beta_{\lambda(nu)n}\lambda(nu)_{ni}+\varepsilon_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+\beta_{\lambda(p)o}\lambda(p)_{oi}+\beta_{\lambda(nu)o}\lambda(nu)_{oi}+\varepsilon_{oi}$	$\lambda(p)_n$ statistically significant and negative, $\lambda(nu)_n$ not statistically significant, $\lambda(p)_o$ statistically significant and negative, $\lambda(nu)_o$ statistically significant and positive
Trinomial logit selection model	Trinomial logit model capturing participation and censored occupation selection decision, and separate wage equations for nurses and all other workers	Workers and non-workers (trinomial logit), workers only (wage equation)	Trinomial logit (selection equation), OLS (wage equation)	$\ln W_{ni}=\beta_n X_{ni}+\beta_{\lambda n}\lambda_{ni}+\eta_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+\beta_{\lambda o}\lambda_{oi}+\eta_{oi}$	λ_n and λ_o statistically significant and negative
Quadrinomial logit selection model	Quadrinomial logit model capturing participation and uncensored occupation selection decision, and separate wage equations for nurses and all other workers	Workers and non-workers (quadrinomial logit), workers only (wage equation)	Quadrinomial logit (selection equation), OLS (wage equation)	$\ln W_{ni}=\beta_n X_{ni}+\beta_{\lambda n}\lambda_{ni}+\eta_{ni}$ $\ln W_{oi}=\beta_o X_{oi}+\beta_{\lambda o}\lambda_{oi}+\eta_{oi}$	λ_n not statistically significant, λ_o statistically significant and negative

[1] The correlation co-efficient ρ_v measuring the correlation between the disturbances in the participation and occupation selection equations is not statistically significantly different from zero. It is therefore appropriate to estimate the two selection equations separately as independent probit models.

[2] X_j = years of full-time education, years of full-time education squared, as a nursing qualification, highest qualification is a postgraduate degree, highest qualification is a first degree, highest qualification is A level, has no qualifications, years of experience with current employer, years of experience with current employer squared, has health problems affect paid work, non-white ethnic group, non-British nationality, non-white and non-British, lives in the South East of England, total usual hours worked per week, employed as a supervisor, manager or foreman, 25 or more workers at workplace, job is non-permanent or temporary, plus 31 time trend dummy variables

Table 7.2. Main features of the models in Chapter 6

There are four main outcomes of the analysis. First, we identified the factors in the wage equation that affect nurses' earnings. Generally the effects of the variables were of the same statistical significance, sign and magnitude as in the single selectivity models though there are some differences. The outcomes pertaining to the selection bias correction terms are presented in Table 7.2. The outcomes in terms of the other co-efficients (the personal and labour market characteristics) are presented in Table 7.3. There is generally much agreement across the different models in terms of the statistical significance, sign and order of magnitude of the co-efficients. This particularly applies to the years of education variables, the work experience variables, the regional variables and the job characteristic variables. Some differences across models occur in terms of statistical significance, sign and order of magnitude of the co-efficients on NURSEQUA, ALEVEL, NOQUAL and DISABLE. Partly this is explained by the interpretation of the co-efficients in Model 5 in Chapter 5 (the occupation selection bias model), where the marginal effects rather than the co-efficients reflect more closely the co-efficients in the other models. Additionally, as pointed out in Chapter 6 the results of the trinomial logit and quadrinomial logit selection models should be treated with caution since they violate the IIA assumption. The main point is that across the models the results are generally consistent, emphasising the plausibility and robustness of the findings.

	Model 3 [1]	Model 4 [1]	Model 5 [1]	Bivariate probit selection model [2]	Independent probit with censoring [2]	Trinomial logit selection model [2]	Quadrinomial logit selection model [2]
Years of education variables							
YED	0.0975*	0.0979*	0.1100*	0.1052*	0.0886*	0.1020*	0.1127*
YED2	-0.0030*	-0.0031*	-0.0035*	-0.0033*	-0.0027*	-0.0032*	-0.0036*
Educational attainment variables							
NURSEQUA	0.2389*	0.2388*	0.6542*	0.2609*	-0.1074	0.2298*	0.2200*
PGDEG	0.2252*	0.2257*	0.1587*	0.1863*	0.2239*	0.1858*	0.2368*
DEG	0.0998*	0.1000*	0.0920*	0.08718*	0.0671*	0.0814*	0.1113*
ALEVEL	-0.1683*	-0.1681*	-0.1415*	-0.0584	-0.0641	-0.0617	0.1510
NOQUAL	-0.2666*	-0.2665*	-0.3274*	-0.0844	-0.0790	-0.1667	-0.0010
Work experience variables							
EXP	0.0197*	0.0197*	0.0194*	0.0208*	0.0204*	0.0206*	0.0208*
EXP2	-0.0003*	-0.0003*	-0.0003*	-0.0004*	-0.0004*	-0.0004*	-0.0004*
Personal characteristic variables							
DISABLE	-0.0435#	-0.0468	-0.0612*	0.0238	0.1396*	0.0712	-0.0571*
ETHNIC	-0.0262	-0.0261	-0.0073	0.0324	0.0488	0.0433	0.0052
NONBRIT	0.0659*	0.0660*	0.0967*	0.0843*	0.0858*	0.0933*	0.0636*
ETHNBRIT	0.0005	0.0004	0.0079	0.0029	0.0101	0.0110	-0.0053
Regional variables							
SEAST	0.0707*	0.0704*	0.0592*	0.0653*	0.0699*	0.0650*	0.0668*
Job characteristic variables							
HOURSPW	-0.0043*	-0.0043*	-0.0042*	-0.0037*	-0.0038*	-0.0038*	-0.0037*
MANAGE	0.0670*	0.0670*	0.0672*	0.0526*	0.0516*	0.0522*	0.0525*
NWORKERS	-0.0402*	-0.0400*	-0.0363*	-0.0377*	-0.0375*	-0.0363*	-0.0361*
TEMP	-0.0790*	-0.0791*	-0.0785*	-0.0749*	-0.0755*	-0.0759*	-0.0785*

[1] From Chapter 5
[2] From Chapter 6
* Significant at the 5% level

Table 7.3. Personal and labour market characteristics affecting nurses' earnings: a synthesis of the results of Chapters 5 and 6

The second main finding concerns the importance of assuming independent disturbances in the multinomial logit model. From the point of view of statistical estimation it is useful that the odds ratios estimated by multinomial logit do not depend on the other choices. However, from a behavioural point of view the IIA assumption is not very attractive. We test for the validity of the assumption in both the trinomial and quadrinomial logit models using a Hausman test (Hausman and McFadden, 1984) and the Small-Hsiao test (Small and Hsiao, 1985). We find that the IIA assumption is violated in both models. In terms of model selection the correct interpretation is therefore that on statistical grounds there is a clear case for preferring the results of the bivariate probit selection models over and above those of the multinomial logit selection models. This result confirms the importance of testing the IIA assumption when interpreting the result of multinomial logit models.

The third main finding concerns the results of the decomposition analysis. Having removed the (positive) endowment component from the earnings differential we find from the results of the bivariate probit selection models that the premium to being employed as a nurse is positive whereas from the multinomial logit selection models the premium is negative. Because the multinomial logit models violate the IIA assumption the results should be treated with scepticism and the results generated by the bivariate probit selection models should be preferred. The upshot is that the wage premium to being employed as a nurse is positive. This means that the average nurse would earn lower wages if paid according to the pay structure of other workers. After controlling for differences in individual and productive characteristics and selection bias nurses are paid higher wages than other workers.

This finding is in direct contrast with the findings from the single selectivity models presented in Chapter 5 (a summary of decomposition results across the two chapters is presented in Table 7.4). It is important to remember that the double selectivity models utilised in Chapter 6 are superior to the more straightforward models in Chapter 5 because they take inot account more fully the economic models of labour market behaviour discussed in Chapter 4. Specifically, according to the two economic models of individual labour supply and occupation selection which provide the economic framework driving the statistical selection bias models one would expect a priori there to be potential selection bias effects arising from an individuals' participation and occupation selection decisions. Only in the double selectivity models are these economic models fully taken into account. Thus from an economic perspective the double selectivity models are to be preferred. This raises concerns for the vast majority of studies conducted in this area (see the literature review in Chapter 4), which usually concentrate on the more straightforward 'single selectivity' framework correcting mainly for potential participation selection bias only.

	Effect on observed difference in mean lnW [1]
Differences in endowments	
OLS model	Positive
Participation selection bias model	Positive
Occupation selection bias model	Positive
Bivariate probit selection model	Positive
Independent probit selection model with censoring	Positive [2]
Trinomial logit selection model	Positive
Quadrinomial logit selection model	Positive
Premium	
OLS model	Positive
Participation selection bias model	Negative
Occupation selection bias model	Negative
Bivariate probit selection model	Positive
Independent probit selection model with censoring	Positive
Trinomial logit selection model	Negative
Quadrinomial logit selection model	Negative
Differences due to selection bias	
Participation selection bias model	Positive
Occupation selection bias model	Positive
Bivariate probit selection model	Positive
Independent probit selection model with censoring	Positive
Trinomial logit selection model	Positive
Quadrinomial logit selection model	Positive

[1] The observed difference in mean lnW is positive

[2] When the premium to being a nurse is analysed using characteristics of non-nurses the differences in endowments is negative

Table 7.4. Results of the decomposition analysis

The fourth main finding concerns the estimation procedure for the bivariate probit selection models. The correct a priori starting point is that the participation and occupation selection decisions may be dependent (made simultaneously). The two equations are estimated by bivariate probit and the correlation co-efficient ρ_v measuring the correlation between the disturbances in the two equations is examined. If ρ_v is statistically significantly different from zero the correct procedure is to estimate the two equations simultaneously in a bivariate probit model. If on the other hand ρ_v is not statistically significant it is appropriate to estimate the two selection equations separately as independent probit models. On this basis for the uncensored model the bivariate probit framework is correct. For the censored model however the independent probit model is more appropriate. If the participation and occupation selection equations in the uncensored model are instead estimated (incorrectly) as two independent probits the relative importance of the selection bias variables is altered. For example, in the rejected independent probit model (see Appendix 6.9 Table A.6.9.2) both the selection bias variables for nurses are statistically significantly different from zero, which is not the case in the bivariate probit model. Additionally a different decomposition is obtained (the differences due to participation and occupation selection bias is negative – see Appendix 6.12). This raises doubts concerning the results of previous double selectivity studies (very small in number) where the independence of the two selection decisions is simply assumed and not tested (see for example Shields and Wheatley Price, 1998, and Botelho et al., 1998).

Taken together the analyses of Chapters 1 to 6 provide a comprehensive picture of nurses' earnings in Great Britain. On the basis of the private net present value to becoming a nurse we conclude that there are financial returns to nursing. Using a novel and sophisticated earnings function approach to analysing nurses' relative earnings we find there is a positive wage premium to being employed as a nurse even after adjusting for differences in labour market endowments. The implication is that in terms of relative earnings there are financial benefits to being employed as a qualified nurse in the British NHS. This is the main finding of the thesis.

Bearing these points in mind we now discuss some policy implications of the analysis. We first relate the finding of a positive wage premium to being employed as a nurse to the process by which nurses' pay is determined as described in Chapter 1 and the bargaining strategy of the Staff Side and the Management Side in pay negotiations. We then discuss some suggestions for reducing the current nursing shortage in light of the findings of the thesis.

7.3. Policy implications

7.3.1. Nurses' pay determination and the bargaining strategies of the Staff Side and the Management Side

In Chapter 1 it was noted that the Pay Review Body takes into account a number of issues in its deliberations concerning nurses' pay: affordability of potential pay rises; recruitment and retention; fairness and comparability; morale and motivation; and, productivity and workload. Using these criteria we found that the Pay Review Body bases its recommendations on nurses' pay on evidence submitted from three main sources: nurses and their representatives (the Staff Side); managers and employers and their representatives (the Management Side); and, the wider economy in terms of labour market conditions (the rate of inflation, average earnings in the economy, and pay settlements in other occupations).

The first implication of the findings is that one could argue that the Pay Review Body has been successful in achieving 'comparable' levels of pay for nurses, which is one of the criteria it states it takes into account when making its recommendations. We found evidence in Chapter 1 to suggest that in setting nurses' pay affordability is an important consideration to the Pay Review Body but that the picture appears to be changing and that the issues of recruitment and retention are becoming more prominent. The results of this thesis suggest that while these may be important issues in its deliberations, the Pay Review Body has also achieved comparability in nurses' pay (depending on how this is defined). The existence of a small positive premium to nurses suggests that their pay is at least comparable to the levels of pay achieved by similar workers. The implication is that if comparability may be interpreted to mean earning the same wages as individuals with the same characteristics then the Pay Review Body has been successful in this regard.

In addition to offering some clues as to the success of the Pay Review Body in its deliberations the results of this thesis also have implications for the bargaining strategies of the Staff Side and the Management Side. Under the broad structure by which nurses' pay is determined we found in Chapter 1 that the Staff Side in their evidence typically emphasise the need for fair pay for nurses, recognising nurses' training and qualifications and their roles and responsibilities in the provision of high quality health care. They also point out frequently that nurses' are paid lower wages than workers in other occupations. The results of this thesis suggest that when differences in individual and labour market characteristics are taken into account nurses' are paid higher wages than workers in other occupations. The premium to being employed as a nurse is positive when the endowment component of the earnings differential is removed. It should be borne in mind that any analysis of relative earnings is fraught with difficulties in terms of delineating the time period covered, the selection of comparator groups, and the definitions of 'earnings'. The interpretation of results will always to a certain extent subjective. Nonetheless the main finding in this thesis is that nurses' wages are at the least comparable with those of workers in other occupations. The implication for the Staff Side in pay negotiations is that, caveats above notwithstanding, it would be sensible to concentrate on demonstrating that nurses' deserve substantial pay rises by additional means. One suggestion is that the Staff Side could concentrate on valuing nurses' output to employers and society as a measure for determining 'fair' pay levels. For example, are nurses' being paid in line with the value of the services they provide? More formally, are nurses' being paid according to their marginal value product (MVP)? For the reasons discussed in Chapter 1 it is not straightforward to define the MVP to employers from employing additional nurses, nonetheless if achievable this would provide an additional source of evidence on fairness in nurses' pay in addition to relative pay arguments (see below).

We also found in Chapter 1 that the Management Side usually do not submit evidence on pay comparability and have warned against placing too much emphasis on comparisons with other employee groups because there was a risk of "cherry picking" comparators. Instead they typically emphasise the need for pay levels that allow them to employ an adequate number of nursing staff within their limited budgets. While the Management Side's sentiments may be correct – i.e. that there are difficulties in making comparison of earnings across occupation groups for the reasons discussed in the previous paragraph – the main finding in the thesis that nurses' are not paid less than other workers when differences in individual and labour market characteristics are controlled for would add credence to the Management Side's bargaining strategy and their preference that nurses' should not receive large increases in pay in order to catch up with earnings of workers in other occupations. This would seem to suggest that the Management Side should use relative pay arguments of the kind developed in this thesis to argue their case in pay negotiations.

7.3.2. Reducing the nursing shortage

We noted at in Chapter 1 that recent estimates place the qualified nursing shortage at around 15,000 (Hancock, 1999). Further, between 1987 and 1995 intakes to nurse training fell between 19,600 and 14,200 per annum (Seccombe and Smith, 1997), while an investigation of the 1991 census showed that only 68% of those of working age with nursing qualifications in England were actually working in the profession (OPCS, 1995). The remainder were split between working in another occupation (16%) and out of paid work (16%; OPCS, 1995). The proportion of leavers in the NHS stands currently at around 14% for qualified nursing staff (see Table 1.2 in Chapter 1), but is higher for nurses who have recently completed their training (Gray and Phillips, 1996, Seccombe and Smith, 1997). Recent evidence also suggests that around 40% of nurses are expecting to leave the profession in the next three years (Beishon, Virdee and Hagell, 1995). Additionally, the fact that turnover is highest for nurses under 35 years of age is an important economic issue since the average cost to the taxpayer to train a nurse is around £50,000 (Shields and Ward, 2001). Further, it has been estimated to cost as much as £5,000 for an NHS Hospital Trust to replace a qualified nurse (Gray and Normand, 1990, Buchan and Seccombe, 1991). The extent of the nursing recruitment and retention problems are therefore considerable (Shields and Ward, 2001).

In general terms there are two broad strategies, which are not mutually exclusive, for reducing a shortage of labour. Wages might be increased in order to attract more workers into the labour market or to stop incumbent workers from leaving. Alternatively non-wage factors associated with a particular occupation might be improved. This second option is important because if an individual associates a particular occupation with non-wage advantages or disadvantages then the financial rewards required to attract the individual into the occupation or keep them in the occupation will be different than if the non-wage factors were negligible or could be ignored. The desirability of the two broad options should be determined by their relative costs and benefits.

Translating this to the nursing labour market implies there are two broad options for reducing the shortage. First, nurses' wages could be increased. This would increase the financial returns to working in the occupation. It would also have the effect of offsetting any non-wage disadvantages that might be perceived. As noted by the Pay Review Body in evidence submitted by the Royal College of Nursing (RCN) "there is a clear relationship between pay and morale." Pay is "a central part of the complex mix of elements affecting morale" and is "the most powerful and most immediate indicator of worth, communicating to existing and potential nurses and to society, how much the profession was valued." (Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 2001). The RCN's membership survey reported in the 2001 Pay Review Body report revealed that only 12% of nurses thought they were well paid for the work they did. 90% felt poorly paid in relation to other professional groups, while 73% felt they could be paid more if they left nursing.

The second option is to reduce the non-wage disadvantages of becoming a nurse. A 1995 survey of 2,483 nurses, midwives and health visitors in Great Britain who were out of service but had returned to the profession were asked what item would make the greatest difference in encouraging or enabling them to return to the profession (OPCS, 1995). The results are presented in Table 7.5. The most important factor was "Greater availability of part-time work, more flexible working hours or job sharing", selected by 14% of respondents. Other popular responses included "Refresher courses including updating in recent developments", "Less bureaucracy and more contact with patients" and "Opportunities to acquaint or reacquaint yourself with nursing or health visiting before making a long term commitment". "Better pay" was the next most important factor, chosen by 6% of respondents. This gives an insight into nature of the non-wage factors contributing to the shortage of nurses.

Most important item	% [1]
Greater availability of part-time work, more flexible working hours or job sharing	14
Refresher courses including updating in recent developments	13
Less bureaucracy and more contact with clients/patients	9
Opportunities to acquaint or reacquaint yourself with nursing or health visiting before making a long-term commitment	7
Better pay	6
Provision of creches/day care for young children	5
More opportunities for developing skills including academic study and retraining in different nursing fields	4
Opportunities to attend confidence building and coping courses	3
More sociable working hours	3
Provision of after school childcare for school age children	3
Less stressful working conditions	3
Better career structure and promotion prospects	3
Better resources to do the job	2
Provision of career counselling including advice on job opportunities and application procedures	2
Schemes to keep out-of-service people in touch	2
Higher status and a better image for nurses, midwives and health visitors in the community	2
Extra increments for returning staff in respect of years out of service	2
Career break arrangements include right of return	1
A period of support after taking up employment	1
More contact and support from management	1
More opportunities for experienced staff	1
More recognition and development of specialist staff	1
More time to implement changes made to the profession as a result of legislation	1
Reduced workload	1
More commitment to equal opportunities	1
More opportunities to observe and learn from experienced staff	0
Assistance with housing and travel costs	0
Better physical conditions (e.g. buildings)	0
No item rated as making a difference	10
n	2,483

[1] Indicates % of total respondents who indicated that the item would make the greatest difference in encouraging or enabling the respondent to return to nursing. Respondents were out-of-service qualified nurses and health visitors and qualified nurses and health visitors who have returned to nursing having worked outside the field.
Source: OPCS (1995)

Table 7.5. The item would make the greatest difference in encouraging or enabling the respondent to return to nursing

Additionally, as noted recently by the Pay Review Body: "A thorough examination of the evidence suggests to us that whilst pay may be influential on motivation and morale, it is far from being the over-riding factor. Equally important is the fact that staff feel overworked and under considerable stress. This is often related to a second issue which consistently comes through, which is the difficulty staff find in establishing an acceptable work/life balance, particularly where they have young children." (Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 2001). Further, as noted in the evidence submitted to the Pay Review Body by the NHS Confederation "...while the NHS still had shortages in some staff groups and geographical areas, these were either a symptom of general supply side problems or were caused by factors other than pay. It [the NHS Confederation] considered that a higher overall pay increase was not the solution to either problem."

Using data on 9,625 qualified NHS nurses from a 1994 national survey of NHS nursing staff conducted by the Policy Studies Institute for the Department of Health, Shields and Ward (2001) estimate the factors affecting the determinants of job satisfaction for nurses, including pay and non-wage factors. The analysis is based on a utility (U) function of the following form:

$$U = U(Y, H, RY, IND, JOB, EMP, NURSE, WV) \qquad [7.6]$$

where Y is wages, H is hours worked, RY is relative or comparison wages (based on what individuals in other public sector occupations are likely to earn with the same observable human capital characteristics), IND represent individual specific characteristics (age, sex, marital status, number of dependent children, ethnic group, education), JOB represents job characteristics (nursing grade, specialty, past and present training episodes, job tenure, shift pattern and trade union membership status), EMP represents employer characteristics (type, size and location of NHS employer), NURSE represents aspects of the nursing work environment (being in a shift pattern not equal to the preferred pattern, having control of work hours, participating in unpaid overtime, undertaking work tasks below those expected at each grade, acting up to a higher grade, playing an extended role in the workplace, holding a grade that is an unfair reflection of current nursing duties, being an assessor or mentor of student nurses, working in a workplace where training is encouraged, and having a second job), and finally WV represents variables that capture the importance of pre-determined work values of individual nurses on the decision to enter a career in nursing (helping others, flexible working hours, rewarding work, job security, promotion prospects, and pay).

Based on the results of an ordered probit model the effects of the job characteristics capturing work environment were all found to be statistically significant predictors of job satisfaction, indicating that policies aimed at improving working conditions for nursing would be important in terms of improving recruitment and retention. By far the largest negative determinant of overall job satisfaction was not being graded fairly in accordance with ones duties. Nurses undertaking tasks below their grade, undertaking duties which are typically undertaken by more senior staff or those working unpaid overtime also reported significantly lower levels of job satisfaction. The largest positive effect on job satisfaction originated from being in a workplace where training was encouraged. In terms of the work values those for whom the flexibility of hours and helping others were principal reasons for entering the nursing profession indicated significantly higher levels of job satisfaction. A preference for rewarding work had the largest effect on the probability of reporting to be satisfied with their job. Those nurses emphasising the more pecuniary aspects of the job such as job security, promotion prospects or pay reported lower, although not significantly lower, levels of overall job satisfaction. As expected absolute wages (proxied in this analysis by grade) were positively associated with job satisfaction.

The second part of the analysis confirmed the a priori expectation that job satisfaction is an important consideration in determining intentions to quit NHS nursing. The authors found that nurses who reported overall dissatisfaction with their jobs had a 65% higher probability of intending to quit than those who reported they were satisfied. They also found however that dissatisfaction with promotion and training opportunities were found to have a stronger impact than workload or pay. Shields and Ward (1995) conclude that policies which focus heavily on improving pay will have only limited success in improving recruitment and retention unless they are accompanied by improved promotion and training opportunities.

The upshot from the above discussion is that there are two broad strategies for reducing the nursing shortage. Wages might be increased in order to attract more nurses into the profession or to stop incumbent nurses from leaving. Alternatively/additionally non-wage factors such as promotion and training opportunities, flexible working hours and the provision of child care facilities associated with a career in nursing might be improved. Both these strategies will affect to a greater or lesser extent affect recruitment and retention in the nursing profession. Also both strategies will incur costs in their provision. Enhancing nurses' pay will increase the size of the nursing paybill, and the improvement of non-wage factors will also have cost implications. Which of these two broad options (or combinations of the two options) should be used to reduce the nursing shortage should be determined in an incremental cost-benefit analysis.

What this thesis adds to the discussion of strategies for addressing the nursing shortage is that after controlling for individual and productive characteristics and selection bias nurses on average earn higher wages than other workers. Therefore there are financial returns to being employed as a nurse. Even though this is the case there is still a shortage of nurses. This means that nursing labour supply is not responding to the existence of the wage premium for nurses that exists at current wage levels. With these facts in mind the implication is that to succeed in reducing the nursing shortage via wages, nurses' pay will have to rise so that it is higher than that of workers in other occupations. It is not enough that nurses' wages are *equal* to those of other workers, because as is shown in this thesis this is already the case and yet a shortage remains.

As discussed previously it is a common perception that nurses' are paid lower wages than other workers. As we have shown this is not in fact true. This may suggest that to reduce the shortage one related strategy is to alter the *perception* of nurses' poor relative earnings rather than their actual earnings. It may be that all that is required to sustain the shortage is a perception that nurses are paid lower wages than other workers, regardless of whether or not they actually are paid less. In other words the important issue is not necessarily that nurses are paid higher wages than other workers, but that they are *perceived* to earn lower wages. While there may be actual positive financial returns to becoming a nurse, if potential new entrants and incumbents generally perceive that the returns are negative then this may convince them to choose a career other than nursing or leave the profession. One implication is that to reduce the nursing shortage the existence of a wage premium to being employed as a nurse should be widely publicised and disseminated. This also provides an interesting interpretation of the effect of the Staff Side's strategy in nursing pay negotiations.

7.4. Some shortcomings and issues for further research

The main finding of this thesis is that after controlling for individual and productive characteristics and selection bias nurses on average earn higher wages than other workers. While for the reasons discussed above this is an interesting finding with some important policy implications it is acknowledged that there are a number of limitations. These arise mainly from restrictions in the available data which have limited the scope of the research. We now discuss these in the context of offering some suggestions for further research and we offer some interesting issues that may be developed into future research questions.

7.4.1. Estimating geographical differences in the returns to nursing

As noted in Chapter 1 local labour market demand and supply conditions are likely to affect the prevailing conditions vis-à-vis employment and wages in the nursing labour market, and these are likely to permeate and modify the issues raised throughout the thesis. Even with national pay scales the magnitude of the premium to being employed as a nurse may differ across geographical areas. For example, a consistent finding in Chapters 5 and 6 was that the co-efficient on SEAST (living in the South East of England) was statistically significant and positive for nurses and that it was lower for nurses than for other workers. The interpretation is that while nurses working in the South East of England receive a premium relative to their nursing colleagues working elsewhere this premium is lower for nurses than for other workers. This has implications for potential geographical differences in the nursing shortage. For example, one possibility is that nurses working in the South East of England, particularly London, may not be able to afford to obtain a mortgage to buy a house and this might acts as a disincentive to work as a nurse in London.

The upshot is that it would be useful and informative to analyse differences in the returns to being employed as a nurse relative to other workers across geographical areas. Such an analysis would require detailed information on nurses' earnings and their individual labour market characteristics with geographical identifiers, and with sufficient numbers of nurses in each geographical area to allow a significant comparison. While the appropriate data for each individual is available in the Quarterly Labour Force Survey (QLFS) data utilised in this thesis, the problem currently is that a much larger sample of nurses is required to conduct a meaningful analysis at the local level. It is worth bearing in mind however that the QLFS is an ongoing survey and therefore the size of the dataset will grow over time. The upshot is that one would fully expect to be able to address this issue in the future.

7.4.2.Estimating the returns to different types of nursing

As noted above it would be useful to analyse differences in the returns to being a nurse across geographical areas. It would also be informative to know the premium to different kinds of nursing within the profession. For example, as noted in Table 1.1 in Chapter 1 there are different types of nurses that work in the NHS (acute, elderly and general nurses, paediatric nurses, maternity nurses, psychiatric nurses, learning disability nurses, community nurses and education nurses). Just as local variations in labour market conditions are likely to affect pay and employment across geographical areas, so diverse supply and demand conditions for different types of nurses will cause differences in wages and employment within the profession. This may impact on the relative attractiveness of a career in different parts of the nursing profession.

It would therefore be interesting to compare the returns to different types of nursing. This would help to ascertain whether it was more lucrative to work as, say, an adult nurse rather than a children's nurse. This in turn might shed some light on the existence of shortages within the profession across categories of nurses. The analysis requires the kind of detailed individual-level data used in this thesis. In addition, however, more detailed information is required on occupation status. For example, the QLFS itemises the occupation class of workers (e.g. being employed as a nurse) but does not go into more specific detail in terms of the type of nurse.

7.4.3. Examining whether individuals from poorer family backgrounds are more likely to become nurses

In Chapter 3 we calculated the private NPV and the private IRR to becoming a nurse in Great Britain for the period 1991-1996. Using the NPV investment criterion, we found that nursing is the preferred option for individuals with a high MTPR relative to the market rate of interest (8%-12% or more). Applying the reasons for incorporating discounting in project appraisal we developed the proposition that according to economic theory the MTPR is likely to be higher for individuals from poorer background and therefore such individuals might find a career in nursing an attractive option. This is a testable hypothesis and poses an interesting research question in terms of the factors that influence choosing a career in nursing.

The methodological approach would be to include as variables in an occupation selection equation (of the kind estimated in Chapter 5) additional variables on family background, such as the social class and income of the individual's parents. (As an aside it would also be interesting to see if having a parent who is/was a nurse influences the decision to become a nurse). This requires detailed information on the characteristics not only of the individual but also of their parents and family background, which is not available in the QLFS. It is available in other surveys such as the British Household Panel Survey (BHPS), but unfortunately due to the current size of the survey the number of nurses in the BHPS sample is too small. Like the QLFS the BHPS is an ongoing survey and so it should be possible to address this research question in the future.

7.4.4. Estimating the marginal value product (MVP) of employing an additional nurse

In Chapter 1 we noted that a key feature of the demand for nursing labour was the difficulty in quantifying the marginal value product (MVP) to employers from employing an additional nurse. The MVP is affected by the additional contribution to output (the marginal product of labour, MP_L) and the price of the final good. The major difficulty is that it is not straightforward to measure the contribution of nurses to the production of health care. Difficulties in measuring the MP_L arise because no ideal method exists for measuring the output of the NHS and because it is hard to separate out the specific contribution to output of individual nursing staff.

An interesting research question would therefore be to attempt to quantify the MVP of employing an additional nurse. This would be useful from a policy perspective because it would provide one benchmark against which 'fair' pay for nurses could be determined – if nurses were paid according to their MVP then they are effectively being paid a wage that is equal to the value of their output. This would be of interest to the Staff Side and the Pay Review Body in pay negotiations who both state that fairness is one of their goals in determining nurses' pay.

One approach might be to obtain survey data from hospitals on the number of nurses they employ along with data on the other inputs that go into the provision of health care and use these as independent variables in a regression model with output (measured in some dimension) as the dependent variable. Crudely this would model the change in output resulting from a one-unit change in inputs (e.g. employing an additional nurse). In reality this would require detailed information from hospitals on their inputs and outputs in the provision of health care. There would be significant complicating factors such as adjusting for the impact on quality of the health care provided care as well as the quantity, and clearly the exercise would not be straightforward. Nonetheless it would be interesting and worthwhile from a policy perspective.

7.4.5. Cost-benefit analysis of wage versus non-wage policies in improving recruitment and retention

As noted above there are two broad strategies, which are not mutually exclusive, for reducing the nursing shortage. Nurses' wages might be increased in order to attract more individuals into the profession and to stop current nurses from leaving. Also non-wage factors might be improved. Policies for reducing the nursing shortage should be viewed relative to one another. While increasing nurses' wages may have a positive effect on supply it is also likely to have a substantial financial impact in terms of the nursing paybill. Similarly, improving non-wage factors by, for example, providing training courses, providing crèches and day care for young children and after-school care for school age children, may improve job satisfaction and enhance recruitment and retention, but they are also likely to incur significant costs. The upshot is that it is not possible to determine the relative merit of focusing on non-wage rather than wage factors without more detailed information on the relative costs and benefits of the different policies.

An interesting issue for future research would therefore be to analyse the relative costs and benefits of different strategies for reducing the nursing shortage (increasing wages, improving non-wage factors, or different combinations of the two). As noted in Section 2.3.3 above there does exist some research that has looked at the relative effects of wages and non-wage factors on job satisfaction in the nursing profession and decisions to quit (see for example Shields and Ward, 2001). This could be extended to include potential joiners to the nursing profession and then be used to estimate the wage elasticity of nursing labour supply and non-wage-factor elasticity of nursing labour supply. This would provide information on the potential benefits (in terms of the increase in the number of nurses) of policies to improve wages or non-wage factors. This might then be combined with information on the cost of such policies which would allow a cost-benefit calculation to be conducted on the potential net benefits of different policies. This would not be a straightforward task and would require detailed data on the societal costs and benefits of the different strategies. Nonetheless it would be a useful step forward in determining the most appropriate method of reducing the nursing shortage.

7.5. General conclusions

There were two general aims of this thesis. The first was to delineate the factors that affect nurses' earnings. To meet this aim we conducted a comprehensive analysis of the structure of the nursing labour market to determine the factors that affect nurses' pay determination. Additionally by estimating earnings functions for nurses we defined the variables that influence directly nurses' wages. The second general aim was to examine the nature and magnitude of wage differentials between nurses and other workers to see whether nurses and other workers earn comparable wages when other factors are held constant. The results of the decomposition analysis shows that nurses are paid higher wages than other workers. Partly this reflects their superior labour market endowments. However, even after controlling for these differences it is in fact the case that nurses on average earn higher wages than other workers. The main finding of the thesis is therefore that there are financial returns to being employed as a qualified nurse in the NHS in Great Britain.

In meeting the aims we make an original contribution to the literature in three major respects. First, in Chapter 3 we conduct an analysis of the private internal rate of return and the private net present value to becoming a nurse. This is the first time such an analysis has been conducted for nurses in Great Britain. Second, in Chapters 4 and 5 we estimate earnings functions for nurses with appropriate corrections for selection bias. The methodology employed is fairly standard, but the application to the British nursing labour market here is unique. The third original contribution is provided in Chapter 6. We analyse the factors affecting nurses' earnings within a double selectivity framework. A novel set of four statistical models are constructed to analyse nurses' earnings simultaneously adjusting for the effects of the decision to participate in the labour market and the decision to be employed as a nurse.

APPENDICES TO CHAPTER 3

Appendix 3.1. NPV versus IRR as a measure of the attractiveness of an investment in human capital

The following discussion is based on Thompson (1980), Sugden and Williams (1978), and Pearce and Nash (1981).

There are a number of circumstances when the NPV and IRR rules may yield conflicting results regarding the attractiveness of an investment in human capital and/or when one or both rules may lead to incorrect decisions. The four main areas of potential difficulty are as follows:

1. Mutually exclusive investment opportunities;
2. The absence of a unique IRR and/or the relationship between NPV and marginal time preference rate (MTPR) is not monotonic;
3. Both the IRR and the NPV ignore the time pattern of costs and benefits; and,
4. The MTPR may vary over the lifetime of the project.

A3.1.1. Mutually exclusive investment opportunities

Many human capital investment decisions by the individual involve a choice between mutually exclusive projects. The problem raised by mutually exclusive investment opportunities is that inappropriate specification of the opportunity cost of a decision may lead to discrepancies between the NPV and IRR rules concerning the attractiveness of the investment opportunity. In terms of the NPV rule there are no ambiguities. Faced with a single possible investment in human capital the NPV criterion dictates that the individual should undertake the investment if the NPV exceeds zero. Where the choice is between different investments in human capital and these investments are mutually exclusive the general rule is to select the project with the highest NPV at the marginal time preference (MTPR) selected. In the case of the IRR rule however, complications may arise. Faced with a single possible investment in human capital the IRR approach is to calculate the IRR and compare it to the individual's MTPR. The rule for undertaking the investment is to accept if the IRR is greater than the MTPR. When comparison of mutually exclusive investments is required it is not necessarily the case that the investment with the highest IRR should be preferred. Figure A3.1.1 plots the NPV of two hypothetical mutually exclusive projects (A and B) by possible values of the individual's MTPR. The IRR for each project (r_A and r_B) occurs at the MTPR where the NPV is equal to zero. Ranking investments by IRR, Project A would be preferred to Project B. However, the NPV of A is only greater than the NPV of B when the individual's MTPR is greater than the crossover rate r_C. If the individual's MTPR is in fact less than r_C then Project B should be preferred and choosing the investment with the highest IRR is inappropriate. When comparison of mutually exclusive investments is required, the investment with the highest IRR should only be preferred if there is no crossover MTPR. Otherwise the NPV approach should be preferred. It is possible to devise an incremental yield approach to overcome this problem but "the effort is unnecessary since the NPV will give the correct answer with far less effort" (Pearce and Nash, 1981, p. 53). In this chapter, both the IRR and NPV are calculated and compared and assessed. We also plot the NPV by MTPR to deduce the appropriate specification of the opportunity cost of a decision.

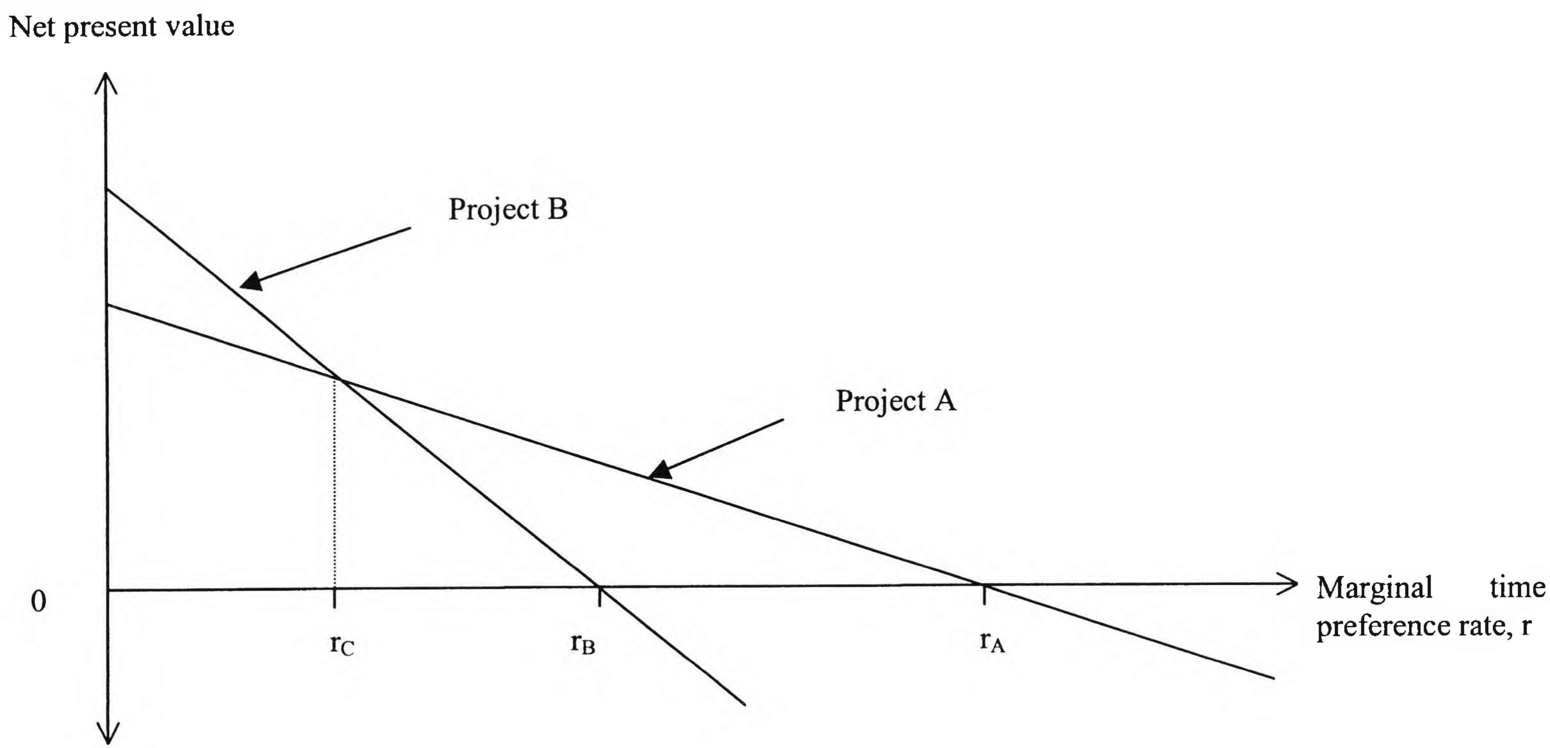


Figure A3.1.1. NPV versus IRR in the comparison of mutually exclusive projects

A3.1.2. The absence of a unique IRR and/or the relationship between NPV and MTPR is not monotonic

The second problem arises because of irregularities in the stream of costs and benefits associated with a project due to more than one sign change in the NPV stream by MTPR. This might arise from a stream in which there were net benefits over the early period of lifetime earnings followed by net costs and then net benefits again (i.e., the NPV is positive then negative than goes back to positive again as the MTPR increases – see Figure A3.1.2). At a low MTPR (up to r^1) the distant in time net benefits together with the early net benefits outweigh the net costs so that the NPV is positive. In the middle range of MTPRs (r^1 to r^2) the early in time net benefits plus the now more heavily discounted time-distant net benefits do not outweigh the net costs and so the NPV is negative. Finally, at a very high MTPR (r^2 and above) both the now heavily discounted net costs and the distant in time net benefits are outweighed by the early net benefits and the net benefits in total outweigh the net costs. In this case the NPV is again positive.

In this situation, as shown in Figure A3.1.2 there are therefore two IRRs. However, this need not necessarily be the case. There may be more than two, depending on the nature of the stream of costs and benefits over the individuals lifetime. Alternatively, if the NPV function were to shift upwards it is possible that there would be no IRR at all. This would imply that the NPV is positive at all MTPRs. Clearly in these circumstances the IRR is inappropriate as a measure of the attractiveness of an investment project.

Note that while there may be no, one or more than one IRR the problem does not apply to the NPV rule because at any specified MTPR there will be only one NPV score. However, it is clear that care must be taken when there is uncertainty concerning the correct MTPR to use. This is because it cannot be assumed automatically that just because the NPV is positive at one MTPR it will necessarily be positive and larger at all lower MTPRs and either positive and lower or negative at all higher MTPRs.

In the present analysis we do plot the NPV by MTPR to determine the uniqueness of the IRR score obtained.

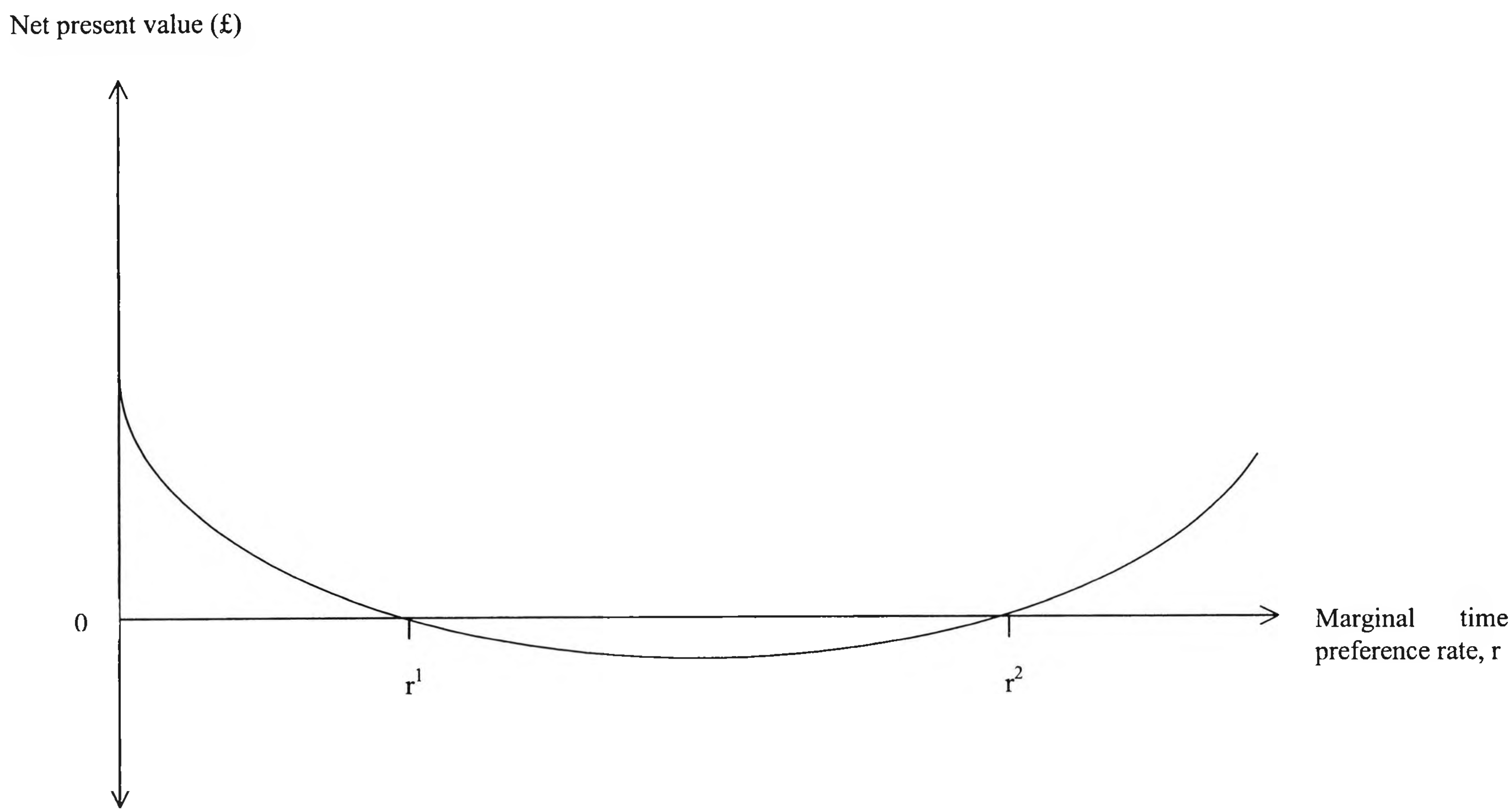


Figure A3.1.2. Non-monotonic relationship between NPV and MTPR

A3.1.2. Both the IRR and the NPV ignore the time pattern of costs and benefits

A further potential problem arises in applying both the NPV and IRR criteria to human capital investments because neither measure tells us anything about the time stream of net benefits. For example, both criteria are insensitive to whether the IRR or the NPV calculated at a specific MTPR is the result of large net benefits in the distant future or small net benefits in the near future. Similarly they are unable to distinguish whether the net benefits consist of a large sum at a single point in time or smaller sums over more than one point in time. However, the individual may be sensitive to this.

One situation that might arise in the comparison of two mutually exclusive projects is that the same NPV may be obtained at the individual's MTPR for both projects, though the individual may not in fact be indifferent between the two projects. This will depend on their preferences for consumption across time periods. Following this line of reasoning, at any given MTPR a project with a high NPV may conceivably not be preferred to one with a lower NPV. This will depend on the nature of the distribution of net benefits over time, the individual's preferences for these, and the various lending and borrowing possibilities available to the individual (for borrowing to consume in earlier time periods and lending to consume in later time periods).

We can only take into account the effect of differences in the time pattern of costs and benefits across projects if we know the individual's consumption preferences across time periods and the individual's borrowing and lending opportunities at each time period.

A3.1.4. The MTPR may vary over the lifetime of the project

The fourth and final problem considered here is that the relevant MTPR for comparisons may vary over the lifetime of the project. On the face of it this does seem to be quite likely given that time preferences may vary with age and the fact that borrowing and lending opportunities vary over the life cycle.

Unfortunately this issue cannot be accounted for using the IRR method, which estimates the constant MTPR required to achieve a zero NPV over the lifetime of the project relative to the opportunity cost option. In the case of the NPV criterion conceptually this problem can be addressed by applying a different MTPRs across the lifetime of the project and discounting each element in the stream of net costs and benefits at the MTPR applicable to the time period of that net cost or benefit. In practice however it is unlikely that the information required for such a comprehensive exercise will be available (indeed it is unlikely that any information on individuals' MTPRs is available).

In the present analysis we calculate the NPV of different human capital investments at different constant MTPRs, ranging from 0% to 40%. While this pragmatic approach does not fully address this issue it does at least allow the attractiveness of investments to be assessed across a range of MTPRs.

A3.1.5. Conclusions

Occasions when the NPV and IRR rules may yield conflicting results and differences between the two investment criteria have been carefully explored. The general conclusion is that both the IRR rule and the NPV rule have a number of potentially serious weaknesses as summary measures of the attractiveness of investments in human capital. However, because of its superiority in comparing mutually exclusive investment opportunities we conclude that on balance the NPV has fewer weaknesses. Nonetheless, the relative ease of calculation of the IRR criterion means that estimation of this measure is included in the present analysis. However, clearly care must be taken when interpreting the results of both the IRR and NPV analyses.

Appendix 3.2. Practical issues in the empirical estimation of NPVs and IRRs

In this appendix we discuss the practical difficulties that arise in calculating the NPV and IRR of investments in human capital. We focus on problems that might arise from a lack of reliable data with which to estimate the costs and benefits of investments and also on fundamental problems associated with the practical approach to estimation. The issues we consider are as follows:

1. Conflation of age and cohort effects in the use of cross-section data;
2. Effects of inflation;
3. Comparative advantage and ability bias;
4. Adjustments for mortality;
5. Adjustments for unemployment;
6. Adjustments for other causes of economic inactivity;
7. Adjustments for discontinuation from training;
8. Non-wage costs and benefits; and,
9. Effects of investments in human capital on work hours.

A3.2.1. Conflation of age and cohort effects in the use of cross-section data

Two types of data could be used in NPV and IRR studies: cross-section data; and, cohort (or longitudinal) data. Clearly cohort data is the most accurate source of data, in the sense that it records actual lifetime earnings profiles for individuals. However, for investments that last over the lifetime of the individual (which is assumed to be the case here) such data must be collected over many years. While cohort data is therefore useful for checking up on the relative attractiveness of past investments these data are not informative at all for measuring the attractiveness of potential current investments. For example, suppose an 18 year old female is considering undertaking a career in nursing in year t = 1 (in the year 1996, say). She decides to base her decision on the NPV criterion based on the lifetime earnings of a nurse (whose career spans 42 years from age 18 years to age 60 years). Using cohort data the individual would be unable to make the calculation until year t = 42 (the year 2038) which would be the last year in which the cohort data on earnings of a 60 year old nurse who invested in 1996 would be accrued. Clearly this is not useful as a measure of the attractiveness of an investment in time t = 1. In this way we can see that cohort data is unhelpful in judging ex ante the attractiveness of an investment in human capital.

The method adopted in this study is to use cross-sectional data on mean earnings of individuals at each age and in each occupation/qualification group. This is normal practice in NPV and IRR studies (see for example, Birch and Calvert, 1973; Burstein and Cromwell, 1985; Maglen and Layard, 1970; Metcalf, 1973; Morris and Ziderman, 1971; Mott and Kreling, 1994; Wilkinson, 1966; Wilson, 1980, 1983a, 1983b, 1984, 1985a, 1985b, 1987a, 1987b; Ziderman, 1973). This may potentially conflate age and cohort effects arising from failure to adjust for growth in earnings over time. In the text (see Figure 3.2) we present some evidence of this using 1980 earnings data. Unfortunately, precisely because future earnings are unobservable it is difficult to estimate the true magnitude of this conflation for the time periods considered in the analysis (1991 to 1996).

One possibility is that some allowance should be made for growth in incomes when cross-section data are used in NPV and IRR calculations. For example, assuming an annual growth rate of earnings of 3% means that measuring the annual earnings of an 18 year old nurse in one years time by the magnitude of a 19 year old nurses' earnings now will underestimate the true earnings by a factor of 1.03. More generally, for estimating earnings within a specific occupation, income in an occupation in t years time with a constant annual growth rate in earnings given by g will be greater than present incomes of those t years older by a factor $(1 + g)^t$.

Further problems arise when comparing earnings across occupation groups. If the same constant rate of growth of incomes is applied across occupation groups at every age then relative incomes will remain constant across occupation groups. However, the absolute value of the earnings differential between occupation groups will not. Unfortunately in reality future annual growth rates in income are unknown, they will probably not be constant over time, and they may also differ across occupation groups. Clearly any adjustment to age-earnings profiles for growth in incomes is far from straightforward.

In summary, the problem with adjusting for growth in incomes in cross-section data is that it will be based on unproven assumptions that by their nature will be very imprecise. In the present analysis we make no adjustments to the cross-section data for income growth over time. However, calculating the NPV and IRR in this way is justifiable for the following reasons. First, we wish to explain the current state of play in the British nursing labour market. In order to calculate the private NPV and private IRR to nursing in 1991 to 1996 we would ideally know the value of future earnings at older ages. For example, we wish to know the future lifetime earnings of an 18-year-old nurse who enters the profession in 1996. As explained above such future earnings are unobservable at the present time. In the context of modelling decisions of occupational choice an individual deciding whether to enter into the nursing profession is unable to ascertain what their true future earnings will be. While future earnings are unobservable what is observable is data on current earnings at older ages. It seems reasonable to suppose that individuals will use current earnings at older ages in an occupation as a predictor of future earnings at older ages. Put simply, an 18-year-old nurse, uncertain as to how much they will earn in 40 years time might reasonably look to what a 58-year-old nurse earns now as an indicator.

Second, from equations [3.1] and [3.2] in the text we can see that future costs and benefits are discounted anyway in the NPV and IRR calculations. The NPV is calculated by discounting future costs and benefits by the individual's MTPR. Computationally, the IRR may be thought of as a discount rate by which future earnings differentials are converted into present values. Therefore, the importance of future unobservable earnings differentials diminishes with time. Possible divergence between unobservable earnings at older ages in future years and currently observable earnings at older ages becomes less important to the individual's choice of occupation.

Third, even if current earnings at older ages (cross-section data) under- or overestimated future earnings at older ages (cohort data) then it is unclear what effect this would have on the NPV and IRR calculations because we are interested not in absolute earnings levels but in earnings differentials between occupation groups.

A3.2.2. Effects of inflation

A related issue is the effect of inflation. Estimating age-earnings profiles using cohort data we would ideally adjust earnings in each year/age following the first year/age of the initial investment to constant prices to allow for changes in the purchasing power of incomes over time. It would then appear that when using cross-section data an allowance should be made for growth in *real* incomes over time. However, for cross-section data the additional adjustment for inflation is unnecessary since purchasing power remains constant across the cross-sectional age-earnings profile. The underlying assumption is that the age-earnings profiles reflect expected real incomes by age.

A3.2.3. Comparative advantage and ability bias

The earnings data used in estimating NPVs and IRRs are usually based on mean earnings at different ages. From the point of view of estimating the private NPV and the private IRR this is potentially problematic. If it is the case that individuals self-select into occupations in which they have a comparative advantage in terms of natural ability and motivation then comparing mean earnings of nurses by age to those in other alternative occupations in a model of occupational choice may be unrealistic. The ideal opportunity cost age-earnings profile would depict the earnings the individual would have received had they not made the investment. In this analysis the earnings of female non-manual workers is used to represent the opportunity cost of becoming a nurse or teacher or obtaining a degree. If individuals who undertake investments in their human capital and then enter a specific occupation have a comparative advantage in that occupation then the earnings of non-manual workers may not accurately reflect the opportunity cost earnings of undertaking nurse or teacher training or obtaining a degree – the opportunity cost earnings in this instance may be overestimated and therefore the NPV and IRR calculations are likely to underestimate the true returns. For this reason we also calculate in the present analysis the NPV and IRR using earnings of all female workers and females whose highest academic qualifications are A levels as opportunity cost earnings. We would generally expect on average earnings of non-manual workers to be greater than earnings of average workers, which in turn would be greater than earnings of workers whose highest academic qualifications are A levels, though this might not necessarily be the case at each age across the age-earnings profile. For an individual who is highly motivated in becoming a nurse and with considerable ability in that area but with lower than average ability elsewhere the earnings of all workers or workers whose highest academic qualifications are A levels would be more appropriate as the opportunity cost earnings. If individuals choose to work in occupations in which they have a comparative *dis*advantage then the assumption is that mean earnings of non-manual workers will underestimate the true opportunity cost earnings. In this instance the NPV and IRR will overestimate the true returns. This option is not considered further here since on balance it is unlikely that an individual will choose to work in an occupation where they are disadvantaged in this way.

Potential difficulties similarly arise comparing different investments in human capital. For example, it might not be the case that a nurse would earn the average wage of a teacher if that nurse became a teacher. This comparative advantage effect means that average earnings data may not accurately represent the benefits to an individual of undertaking an alternative investment in human capital. We discuss this problem and other potential solutions to it in greater detail in Chapters 4-6.

A3.2.4. Adjustments for mortality

In most studies of this kind it is usual to adjust age-earnings profiles for mortality. The approach usually adopted, and the one adopted in this study is to multiply the earnings at each age in the occupations being compared by the cumulative probability that the individual will survive to that age. While we include the adjustment here, it is open to debate whether or not individuals making private decisions do in fact allow for the probability of mortality in such a calculated way. Due to data limitations (a lack of data on occupation-specific mortality rates by age) it is usual to apply the same mortality rates across occupation groups. In general terms this might be problematic if those earning lower incomes (generally in lower socio-economic groups) have a shorter life expectancy than those earning higher incomes. However, the adjustment is naturally very small except for older ages at the end of the career, which are in any case heavily discounted.

A3.2.5. Adjustments for unemployment

The returns to training are likely to be affected positively by the higher employment rates of more educated individuals. Indeed, improved employment rates may be one reason why individuals choose to undertake training for specific occupations in the first place: unemployment among the highly educated is low and very low by comparison with the less well educated. On the other hand non-zero unemployment rates even among the educated may impact on the investment decision in that any chance that a human capital investment might potentially be wasted through unemployment might have a bearing on the individual's decision to undertake the investment in the first place.

In some studies an adjustment for unemployment might implicitly be included in data on earnings anyway. For example, annual earnings data for workers might include earnings of individuals working at the time of the survey who spent part of the year unemployed. Additional adjustment for unemployment would then lead to potential double counting problems. This is not an issue in the present study where weekly earnings data for workers only are extrapolated to annual levels.

All in all it would appear that some adjustment to age-earnings profiles is justified for employment. Earnings at each age are multiplied by the probability that individuals of that age are employed. Occupation-specific rates are applied to account for differences in unemployment rates across occupation groups.

A3.2.6. Adjustments for other causes of economic inactivity

For many females an assumption of full participation in the labour market from age 18 years to retirement is unrealistic. Getting married and having children together with consequent breaks in labour market participation are more likely to be the norm. Additionally, it is probably unlikely that participation rates will be independent of the career path chosen because hours and times of work and the possibilities of part-time employment will vary across occupations.

Individuals choose to participate in the labour market if the wage they are offered for employment exceeds their reservation wage. The offered wage will be affected by education, experience and other individual productive characteristics. The reservation wage will be affected by the level of property (i.e. unearned) income – which may or may not include partners' income – and other family variables likely to affect attitudes to work such as the presence of children in the household. We do consider the impact of female labour force participation on earnings in greater detail in Chapters 4-6. However, clearly some adjustment for non-participation in the workforce is warranted in the present analysis since this is likely to affect the NPV and IRR. Age-earnings profiles in each occupation at each age are adjusted for non-participation due to retirement, family care, long term sickness or disability and maternity leave. The approach adopted is to multiply the earnings at each age by the probability that the individual will participate in the workforce at that age. Due to a lack of occupation-specific data it is assumed that participation rates for reasons other than unemployment do not differ across occupation groups.

A3.2.7. Adjustments for discontinuation from training

While an individual may decide to undertake training as an investment in human capital there is no guarantee that they will successfully complete that training/investment. Some allowance is therefore justified for accounting for the possibility of failure to complete the training if the return on investment in human capital of those who drop out is different from the return on investment of those who complete. In the present analysis we adopt a pessimistic approach by assuming that an individual who drops out earns a zero return on the investment and achieves the earnings profile they would have achieved had they not begun training in the first place (they receive the opportunity cost earnings). We apply different discontinuation rates for different investments and replace the earnings profile of a career in nursing (say) with an expected earnings profile (a weighted profile of earnings in nursing with those in the opportunity cost group) that takes into account the small probability of discontinuation and achieving the lower earnings profile of the opportunity cost group.

It should be borne in mind, however, that in the context of private NPV and private IRR calculations it is debatable whether individuals allow for the possibility of failure in this way. This is because the chances of discontinuation are assessed differently by different individuals each of whom has private asymmetric information about their own capabilities (for example, how hard they are prepared to work and how intelligent they perceive themselves to be). Additionally, the consequences of failure on the attractiveness of an investment may also differ across individuals. For example, individuals may make different assessments of the consequences of failure and may have different attitudes towards the risk of failure. However, since the rates of discontinuation are in fact very low anyway this is unlikely to have more than a negligible impact on the calculations.

A3.2.8. Non-wage costs and benefits

As noted in the text, a key feature of NPV and IRR analyses is that while they measure the costs and benefits of investments in human capital they tend to concentrate on quantifiable economic costs and benefits, including the financial cost of training and education and the subsequent financial earnings of the individual. They usually ignore other non-wage factors associated with investments in human capital. These might include non-wage advantages and disadvantages associated with training and education programmes and similar advantages and disadvantages associated with different occupations. Non-wage effects of education and training (for example, undertaking nurse training, or teacher training, or obtaining a degree at university) are those experienced while actually undergoing education and training (for example, enjoying the lifestyle of a student). Non-wage effects of specific occupations might be positive (for example if a particular occupation had agreeable working conditions and/or a high level of job satisfaction) or negative in a generally disagreeable occupation. A potential problem with the NPV and IRR calculations arises if the financial costs and benefits associated with different occupations differ significantly from the non-financial costs and benefits and where including the full effects (financial and non-financial) would lead to a different decision. Unfortunately, without valuing accurately in monetary terms the non-wage costs and benefits of different forms of training and education and different occupations (for example, through stated preferences techniques) it is difficult to include the non-wage effects.

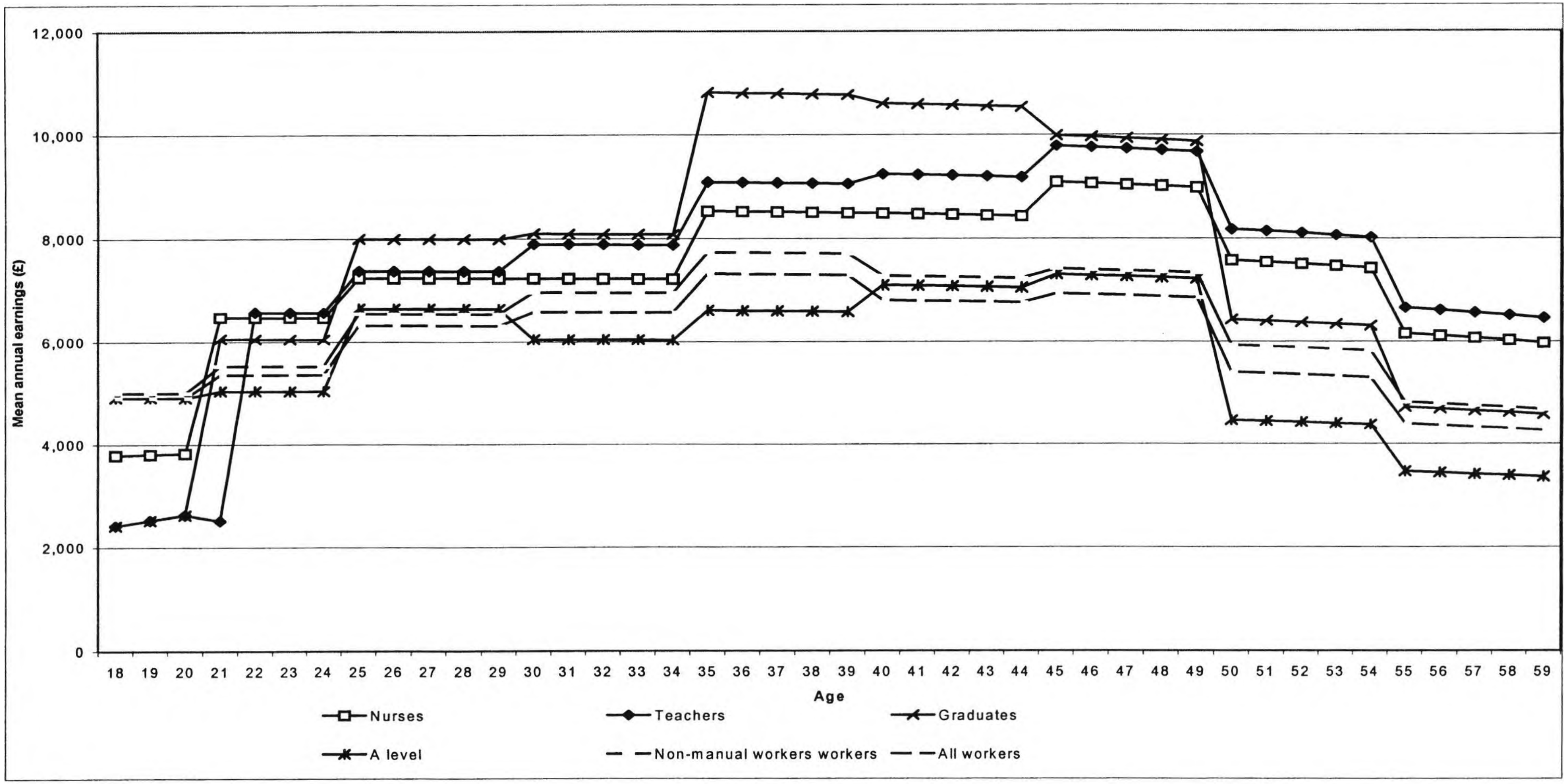
There are two sets of circumstances where it is possible to make some general comments pertaining to the sign of the non-wage effects however. First, if the demand for workers in a specific occupation is greater than the supply (for example, there are job vacancies) and yet the financial rewards are high then ceteris paribus this indicates that the non-wage effects are negative and serious enough to outweigh the financial advantages. Second, where there are unemployed workers (supply is greater than demand) in a specific occupation and the financial rewards are low this suggests ceteris paribus that there are important non-wage benefits to be reaped. Where demand for workers is greater than supply and there are only meagre financial rewards, or where supply is greater than demand and there are significant financial rewards the general sign of the non-wage effects is ambiguous.

A3.2.9. Effects of investments in human capital on work hours

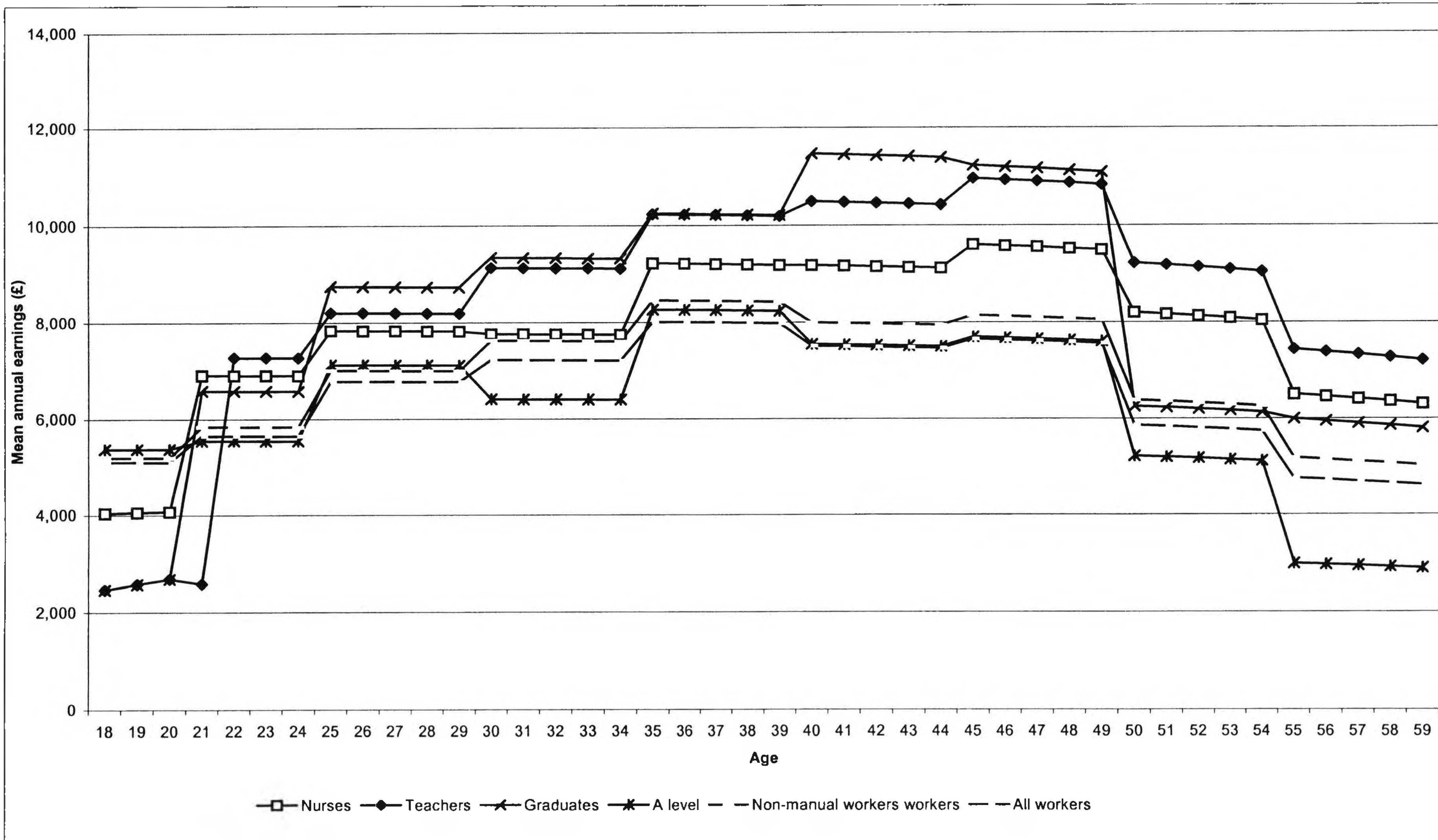
Because workers in some occupation groups work longer hours than others the NPVs and IRRs calculated for these individuals may involve an overstatement of the returns to an investment in human capital. Failure to take account of the longer hours worked means that the measured return may reflect in part a return to working longer hours and not an excess return to investment. As a sensitivity analysis we therefore re-estimate the calculated NPVs and IRRs adjusting for the hours worked in the different occupation groups. All earnings profiles are adjusted to a standard 40-hour working week: earnings at every age in each occupation group are multiplied by a factor of 40/h where h is the mean hours worked per week in that occupation group.

While this adjustment does at least partly address the effects of investments in human capital on work hours there are two additional issues that should be borne in mind. The first is that an adjustment for the number of days/weeks worked per year may also be warranted. Shorter than average holidays together with longer hours worked per week are both likely to lead to an overstatement of the returns to an investment in human capital. Unfortunately, a lack of data on the mean days worked per year in the different occupation groups considered precludes this effect from being factored into the analysis. The second issue is that we are obliged to assume for the purposes of the standardisation process that the mean hours worked per week in each occupation group are constant throughout the working life of the individual. This is due to a lack of data on hours worked per week by age in each occupation. This is probably unrealistic, but is a pragmatic solution given the nature of the available data.

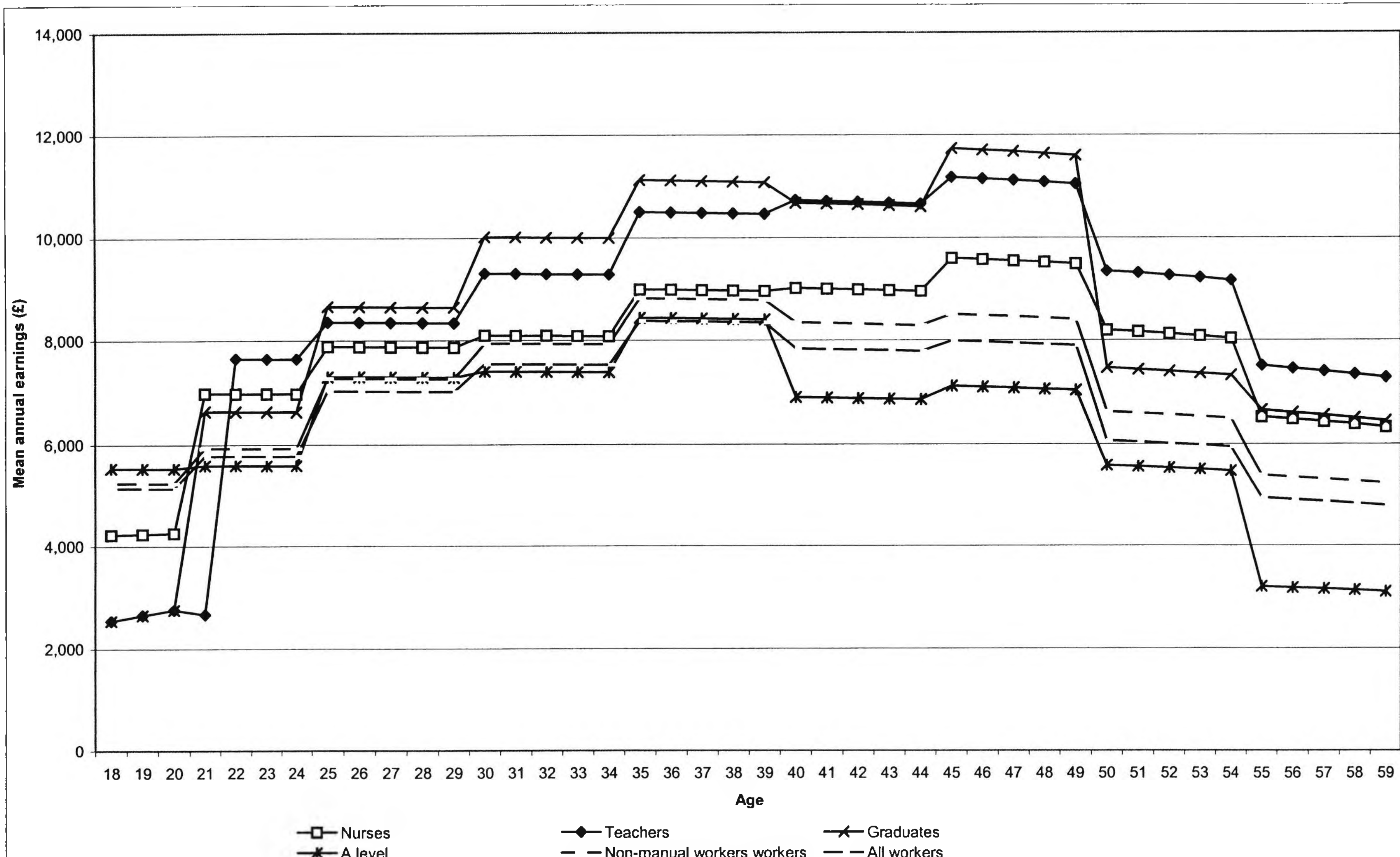
Appendix 3.3. Actual age-earnings profiles used to calculate the private NPV and private IRR for 1991-1995 (earnings net of taxation [income tax plus national insurance] adjusted for mortality, unemployment, other causes of economic inactivity, and discontinuation from training)



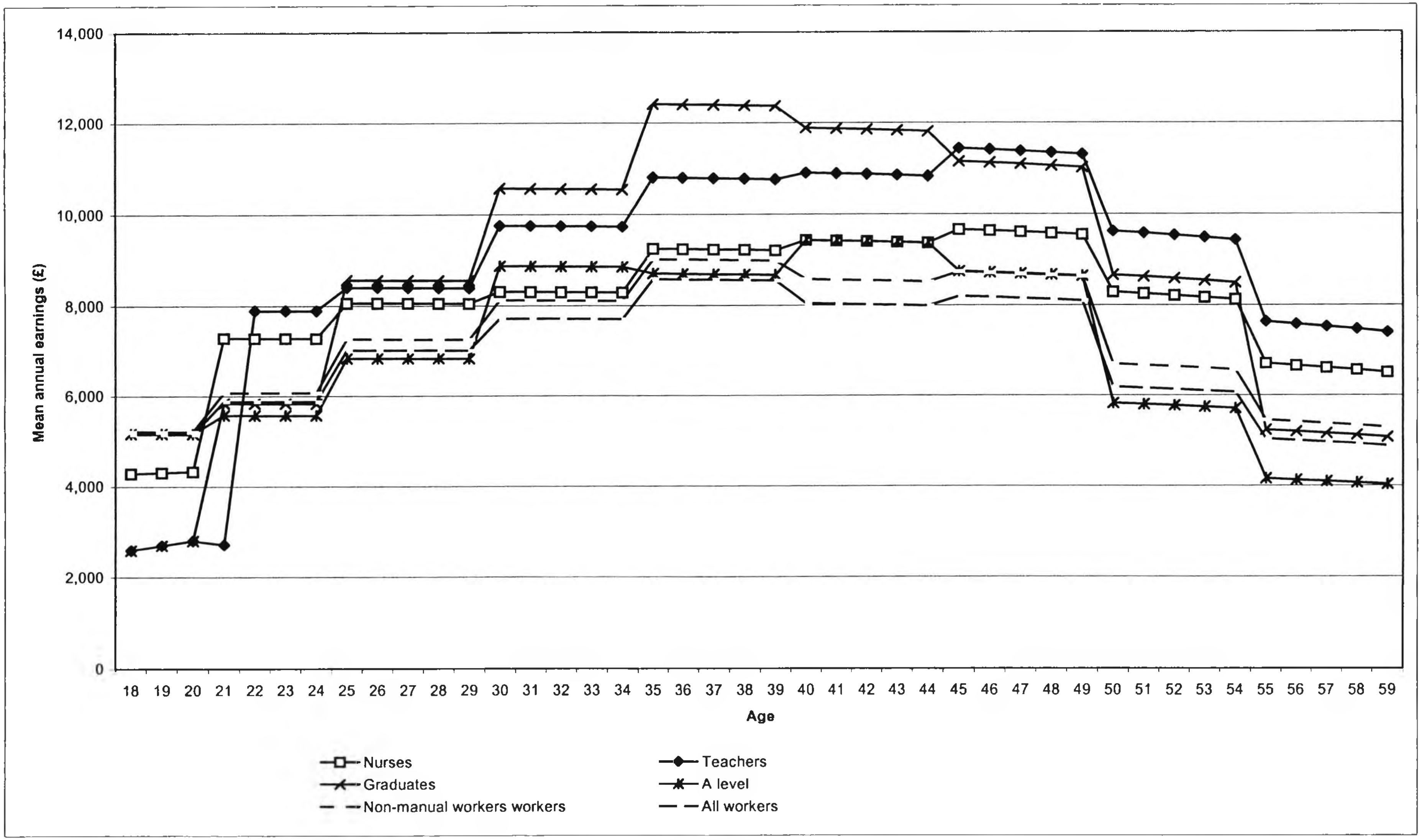
(a) 1991


14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Mean annual earnings (£)
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
Age
Nurses
Teachers
Graduates
A level
Non-manual workers workers
All workers

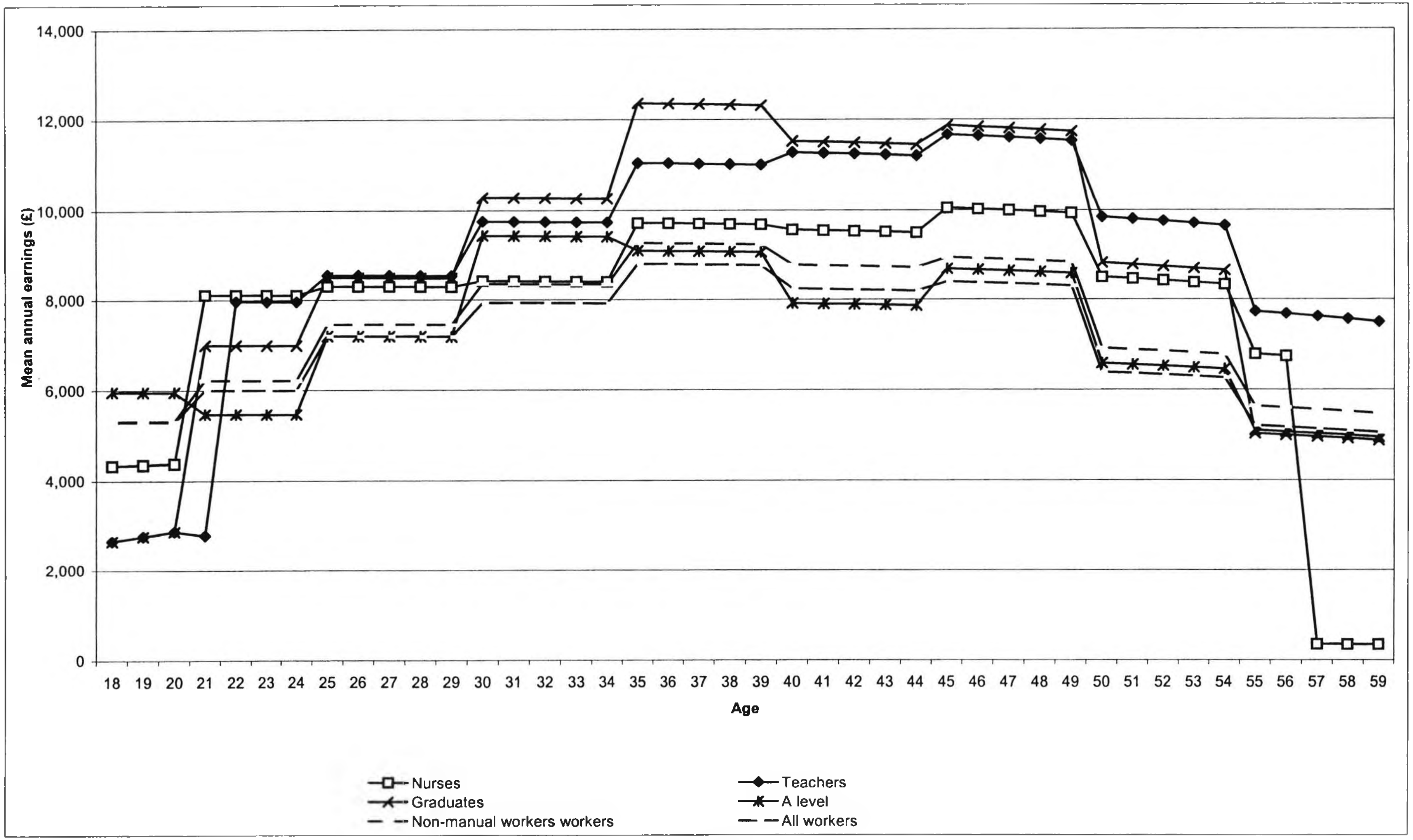
(b) 1992


14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Mean annual earnings (£)
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
Age
Nurses
Teachers
Graduates
A level
Non-manual workers workers
All workers

(c) 1993


14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Mean annual earnings (£)
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
Age
Nurses
Teachers
Graduates
A level
Non-manual workers workers
All workers

(d) 1994



14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Mean annual earnings (£)
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
Age
Nurses
Teachers
Graduates
A level
Non-manual workers workers
All workers

(e) 1995

Appendix 3.4. NPV of nurse training, teacher training and obtaining a degree by MTPR for 1991 to 1996

	Opportunity cost earnings = Non-manual workers			Opportunity cost earnings = All workers			Opportunity cost earnings = workers whose highest qualification is A level		
MTPR	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates
0	38,071	49,950	54,324	53,511	65,390	69,764	65,335	77,214	81,588
1	28,671	36,329	42,573	40,755	48,414	54,658	49,578	57,236	63,480
2	21,791	26,348	33,411	31,379	35,936	42,999	38,077	42,634	49,697
3	16,705	18,969	26,224	24,415	26,679	33,934	29,590	31,855	39,110
4	12,907	13,468	20,556	19,188	19,749	26,837	23,258	23,819	30,907
5	10,042	9,333	16,059	15,224	14,515	21,242	18,480	17,771	24,497
6	7,857	6,199	12,473	12,185	10,527	16,801	14,833	13,175	19,449
7	6,174	3,805	9,598	9,830	7,462	13,254	12,017	9,649	15,442
8	4,863	1,965	7,281	7,986	5,087	10,404	9,819	6,921	12,237
9	3,831	539	5,404	6,526	3,234	8,099	8,083	4,791	9,656
10	3,011	-571	3,877	5,359	1,776	6,224	6,698	3,115	7,563
11	2,353	-1,441	2,629	4,416	622	4,692	5,580	1,786	5,856
12	1,819	-2,125	1,604	3,647	-297	3,432	4,670	726	4,455
13	1,381	-2,665	760	3,014	-1,033	2,392	3,920	-126	3,299
14	1,020	-3,093	61	2,488	-1,625	1,529	3,298	-815	2,339
15	719	-3,431	-518	2,047	-2,103	810	2,776	-1,373	1,540
16	465	-3,699	-1,000	1,674	-2,491	209	2,336	-1,829	870
17	251	-3,911	-1,402	1,357	-2,805	-296	1,961	-2,201	307
18	68	-4,078	-1,739	1,086	-3,060	-721	1,640	-2,506	-167
19	-88	-4,207	-2,020	852	-3,267	-1,080	1,362	-2,757	-569
20	-223	-4,307	-2,256	649	-3,435	-1,384	1,122	-2,962	-911
21	-341	-4,383	-2,454	471	-3,571	-1,642	912	-3,131	-1,201
22	-443	-4,439	-2,619	316	-3,680	-1,860	727	-3,269	-1,449
23	-532	-4,478	-2,758	179	-3,767	-2,047	565	-3,381	-1,661
24	-611	-4,504	-2,874	58	-3,834	-2,205	421	-3,472	-1,842
25	-680	-4,518	-2,970	-49	-3,887	-2,339	293	-3,545	-1,997
26	-741	-4,523	-3,050	-144	-3,926	-2,454	179	-3,602	-2,130
27	-795	-4,520	-3,116	-229	-3,954	-2,551	78	-3,647	-2,244
28	-842	-4,510	-3,170	-305	-3,973	-2,633	-14	-3,681	-2,341
29	-885	-4,495	-3,213	-374	-3,984	-2,702	-96	-3,706	-2,424
30	-922	-4,475	-3,248	-435	-3,987	-2,760	-170	-3,723	-2,495
31	-956	-4,451	-3,274	-490	-3,986	-2,809	-237	-3,732	-2,555
32	-986	-4,424	-3,294	-540	-3,979	-2,849	-298	-3,736	-2,606
33	-1,012	-4,395	-3,309	-585	-3,968	-2,881	-352	-3,736	-2,649
34	-1,035	-4,363	-3,318	-625	-3,953	-2,908	-402	-3,730	-2,685
35	-1,056	-4,330	-3,323	-662	-3,936	-2,929	-448	-3,721	-2,714
36	-1,074	-4,295	-3,324	-695	-3,915	-2,944	-489	-3,709	-2,738
37	-1,090	-4,258	-3,321	-725	-3,893	-2,956	-526	-3,694	-2,757
38	-1,104	-4,221	-3,316	-752	-3,868	-2,964	-561	-3,677	-2,772
39	-1,117	-4,183	-3,309	-776	-3,842	-2,968	-592	-3,658	-2,784
40	-1,128	-4,145	-3,299	-798	-3,815	-2,970	-620	-3,637	-2,791

(a) 1991

MTPR	Opportunity cost earnings = Non-manual workers: Nurses	Teachers	Graduates	Opportunity cost earnings = All workers: Nurses	Teachers	Graduates	Opportunity cost earnings = workers whose highest qualification is A level: Nurses	Teachers	Graduates
0	38,017	67,075	57,206	53,925	82,983	73,114	66,040	95,098	85,230
1	28,753	49,794	44,711	41,216	62,257	57,175	49,468	70,509	65,427
2	21,977	37,061	35,075	31,874	46,957	44,972	37,494	52,577	50,591
3	16,971	27,590	27,584	24,933	35,552	35,546	28,747	39,366	39,360
4	13,231	20,479	21,714	19,720	26,967	28,203	22,286	29,534	30,769
5	10,407	15,090	17,080	15,762	20,445	22,434	17,459	22,142	24,132
6	8,252	10,971	13,396	12,722	15,442	17,866	13,810	16,529	18,954
7	6,587	7,794	10,445	10,363	11,570	14,221	11,019	12,225	14,877
8	5,287	5,323	8,067	8,510	8,546	11,290	8,858	8,894	11,638
9	4,260	3,386	6,138	7,040	6,166	8,917	7,166	6,292	9,044
10	3,440	1,857	4,563	5,860	4,277	6,983	5,826	4,242	6,948
11	2,778	641	3,271	4,903	2,766	5,395	4,751	2,614	5,244
12	2,238	-331	2,204	4,119	1,550	4,085	3,881	1,312	3,847
13	1,792	-1,114	1,319	3,470	564	2,997	3,169	263	2,696
14	1,421	-1,746	582	2,928	-239	2,089	2,580	-587	1,741
15	1,109	-2,259	-35	2,471	-897	1,327	2,088	-1,280	944
16	845	-2,677	-553	2,084	-1,438	685	1,675	-1,847	276
17	619	-3,018	-990	1,752	-1,886	142	1,324	-2,313	-286
18	425	-3,296	-1,359	1,465	-2,256	-319	1,024	-2,697	-761
19	257	-3,523	-1,672	1,217	-2,563	-712	766	-3,014	-1,163
20	110	-3,708	-1,938	1,000	-2,819	-1,048	542	-3,276	-1,506
21	-18	-3,859	-2,164	809	-3,032	-1,337	348	-3,493	-1,798
22	-131	-3,981	-2,356	641	-3,208	-1,584	178	-3,672	-2,047
23	-231	-4,078	-2,520	492	-3,355	-1,797	28	-3,819	-2,261
24	-320	-4,156	-2,660	360	-3,476	-1,980	-103	-3,939	-2,443
25	-399	-4,216	-2,778	241	-3,576	-2,138	-220	-4,037	-2,600
26	-469	-4,262	-2,879	135	-3,657	-2,274	-324	-4,116	-2,734
27	-533	-4,296	-2,964	40	-3,723	-2,392	-416	-4,179	-2,848
28	-589	-4,319	-3,036	-46	-3,776	-2,493	-498	-4,229	-2,946
29	-640	-4,334	-3,097	-123	-3,817	-2,580	-572	-4,266	-3,029
30	-686	-4,341	-3,147	-193	-3,849	-2,655	-638	-4,293	-3,099
31	-727	-4,341	-3,189	-257	-3,871	-2,719	-697	-4,312	-3,159
32	-764	-4,337	-3,222	-314	-3,887	-2,773	-750	-4,323	-3,209
33	-797	-4,327	-3,249	-367	-3,896	-2,819	-798	-4,328	-3,250
34	-827	-4,313	-3,271	-415	-3,900	-2,858	-841	-4,327	-3,284
35	-855	-4,296	-3,286	-458	-3,899	-2,890	-880	-4,321	-3,312
36	-879	-4,275	-3,298	-498	-3,894	-2,917	-915	-4,311	-3,334
37	-901	-4,253	-3,305	-534	-3,885	-2,938	-946	-4,298	-3,350
38	-921	-4,228	-3,309	-567	-3,874	-2,955	-974	-4,281	-3,362
39	-938	-4,201	-3,310	-597	-3,859	-2,968	-999	-4,262	-3,371
40	-954	-4,172	-3,308	-624	-3,842	-2,978	-1,022	-4,240	-3,375

(b) 1992

	Opportunity cost earnings = Non-manual workers			Opportunity cost earnings = All workers			Opportunity cost earnings = workers whose highest qualification is A level		
MTPR	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates
0	28,137	63,868	61,977	44,513	80,243	78,353	63,095	98,825	96,935
1	21,143	47,439	47,830	33,944	60,241	60,631	46,931	73,228	73,619
2	16,079	35,335	37,090	26,221	45,477	47,233	35,286	54,542	56,297
3	12,373	26,331	28,861	20,515	34,474	37,004	26,812	40,771	43,301
4	9,630	19,571	22,498	16,252	26,193	29,120	20,585	30,526	33,453
5	7,576	14,447	17,535	13,029	19,900	22,988	15,960	22,831	25,919
6	6,019	10,529	13,631	10,564	15,074	18,176	12,489	16,999	20,101
7	4,823	7,506	10,536	8,655	11,337	14,367	9,855	12,538	15,568
8	3,893	5,154	8,062	7,159	8,419	11,328	7,835	9,096	12,004
9	3,161	3,309	6,072	5,972	6,120	8,883	6,268	6,416	9,179
10	2,576	1,851	4,459	5,020	4,294	6,902	5,039	4,314	6,922
11	2,103	690	3,143	4,246	2,833	5,286	4,064	2,651	5,104
12	1,716	-240	2,064	3,611	1,655	3,959	3,283	1,327	3,631
13	1,395	-989	1,174	3,083	700	2,862	2,649	266	2,428
14	1,125	-1,595	436	2,640	-80	1,951	2,130	-591	1,441
15	897	-2,089	-178	2,266	-721	1,190	1,700	-1,286	625
16	702	-2,492	-691	1,945	-1,249	552	1,342	-1,852	-52
17	534	-2,821	-1,122	1,669	-1,686	13	1,040	-2,315	-616
18	387	-3,091	-1,484	1,430	-2,049	-442	783	-2,696	-1,089
19	259	-3,313	-1,790	1,221	-2,351	-828	563	-3,009	-1,486
20	146	-3,494	-2,048	1,036	-2,604	-1,157	373	-3,267	-1,821
21	46	-3,643	-2,266	874	-2,815	-1,438	208	-3,480	-2,103
22	-44	-3,764	-2,451	729	-2,991	-1,678	64	-3,656	-2,343
23	-124	-3,861	-2,607	600	-3,138	-1,884	-62	-3,800	-2,546
24	-195	-3,940	-2,739	484	-3,260	-2,060	-174	-3,918	-2,718
25	-260	-4,002	-2,851	380	-3,361	-2,211	-272	-4,014	-2,863
26	-318	-4,050	-2,945	286	-3,445	-2,340	-360	-4,092	-2,987
27	-371	-4,086	-3,024	201	-3,514	-2,451	-439	-4,154	-3,091
28	-419	-4,113	-3,090	124	-3,569	-2,546	-509	-4,202	-3,179
29	-463	-4,130	-3,144	54	-3,614	-2,627	-572	-4,240	-3,253
30	-502	-4,141	-3,189	-9	-3,648	-2,696	-628	-4,267	-3,315
31	-538	-4,145	-3,225	-68	-3,675	-2,755	-679	-4,286	-3,366
32	-571	-4,144	-3,254	-121	-3,694	-2,804	-725	-4,298	-3,408
33	-600	-4,138	-3,276	-170	-3,707	-2,845	-767	-4,304	-3,443
34	-628	-4,127	-3,293	-214	-3,714	-2,879	-804	-4,304	-3,470
35	-652	-4,114	-3,304	-255	-3,717	-2,907	-838	-4,300	-3,491
36	-675	-4,097	-3,312	-293	-3,715	-2,930	-869	-4,292	-3,506
37	-695	-4,078	-3,315	-327	-3,710	-2,947	-897	-4,280	-3,517
38	-714	-4,057	-3,316	-359	-3,702	-2,961	-922	-4,265	-3,524
39	-731	-4,033	-3,313	-388	-3,691	-2,970	-945	-4,247	-3,527
40	-746	-4,009	-3,308	-415	-3,677	-2,977	-965	-4,227	-3,527

(c) 1993

MTPR	Opportunity cost earnings = Non-manual workers			Opportunity cost earnings = All workers			Opportunity cost earnings = workers whose highest qualification is A level		
	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates
0	31,189	67,680	64,314	47,476	83,966	80,600	39,761	76,251	72,885
1	23,788	50,540	49,918	36,542	63,295	62,672	29,849	56,602	55,979
2	18,402	37,896	38,787	28,522	48,016	48,907	22,829	42,322	43,214
3	14,438	28,476	30,124	22,571	36,610	38,258	17,797	31,836	33,484
4	11,483	21,390	23,340	18,102	28,010	29,960	14,145	24,052	26,002
5	9,253	16,009	17,995	14,705	21,462	23,448	11,455	18,212	20,198
6	7,547	11,884	13,760	12,090	16,428	18,303	9,444	13,781	15,657
7	6,224	8,692	10,384	10,053	12,521	14,213	7,916	10,384	12,076
8	5,184	6,202	7,681	8,444	9,462	10,941	6,734	7,752	9,231
9	4,355	4,241	5,505	7,158	7,044	8,308	5,804	5,690	6,953
10	3,686	2,686	3,745	6,118	5,119	6,178	5,058	4,059	5,118
11	3,138	1,443	2,317	5,267	3,573	4,447	4,450	2,756	3,630
12	2,683	443	1,154	4,562	2,322	3,032	3,945	1,705	2,416
13	2,302	-366	203	3,972	1,303	1,872	3,520	851	1,420
14	1,979	-1,025	-577	3,473	469	917	3,156	151	600
15	1,702	-1,564	-1,217	3,047	-219	128	2,840	-427	-79
16	1,463	-2,007	-1,743	2,681	-789	-525	2,563	-907	-643
17	1,254	-2,372	-2,177	2,363	-1,263	-1,068	2,318	-1,308	-1,113
18	1,071	-2,673	-2,534	2,085	-1,659	-1,520	2,099	-1,645	-1,506
19	908	-2,923	-2,828	1,840	-1,991	-1,896	1,902	-1,929	-1,835
20	764	-3,129	-3,069	1,624	-2,269	-2,210	1,723	-2,170	-2,110
21	635	-3,300	-3,267	1,431	-2,503	-2,471	1,560	-2,374	-2,342
22	518	-3,440	-3,429	1,259	-2,700	-2,689	1,412	-2,547	-2,536
23	414	-3,556	-3,560	1,104	-2,866	-2,870	1,275	-2,695	-2,699
24	319	-3,651	-3,665	964	-3,005	-3,020	1,149	-2,820	-2,835
25	233	-3,727	-3,749	838	-3,122	-3,144	1,033	-2,927	-2,949
26	155	-3,789	-3,815	724	-3,220	-3,246	926	-3,018	-3,044
27	83	-3,838	-3,865	620	-3,301	-3,329	826	-3,094	-3,122
28	18	-3,875	-3,903	525	-3,368	-3,396	734	-3,159	-3,186
29	-42	-3,903	-3,929	438	-3,423	-3,449	648	-3,213	-3,239
30	-97	-3,923	-3,946	358	-3,468	-3,491	568	-3,258	-3,281
31	-147	-3,936	-3,955	285	-3,503	-3,522	494	-3,294	-3,314
32	-193	-3,942	-3,957	218	-3,531	-3,545	424	-3,324	-3,339
33	-236	-3,943	-3,953	156	-3,551	-3,561	360	-3,348	-3,357
34	-275	-3,940	-3,944	99	-3,566	-3,570	299	-3,366	-3,370
35	-311	-3,933	-3,932	47	-3,575	-3,573	243	-3,379	-3,377
36	-345	-3,922	-3,915	-2	-3,579	-3,572	190	-3,388	-3,381
37	-375	-3,909	-3,896	-47	-3,580	-3,567	140	-3,393	-3,380
38	-404	-3,893	-3,874	-88	-3,577	-3,558	94	-3,395	-3,376
39	-430	-3,874	-3,850	-127	-3,571	-3,547	51	-3,393	-3,369
40	-454	-3,854	-3,824	-162	-3,562	-3,532	10	-3,389	-3,360

(d) 1994

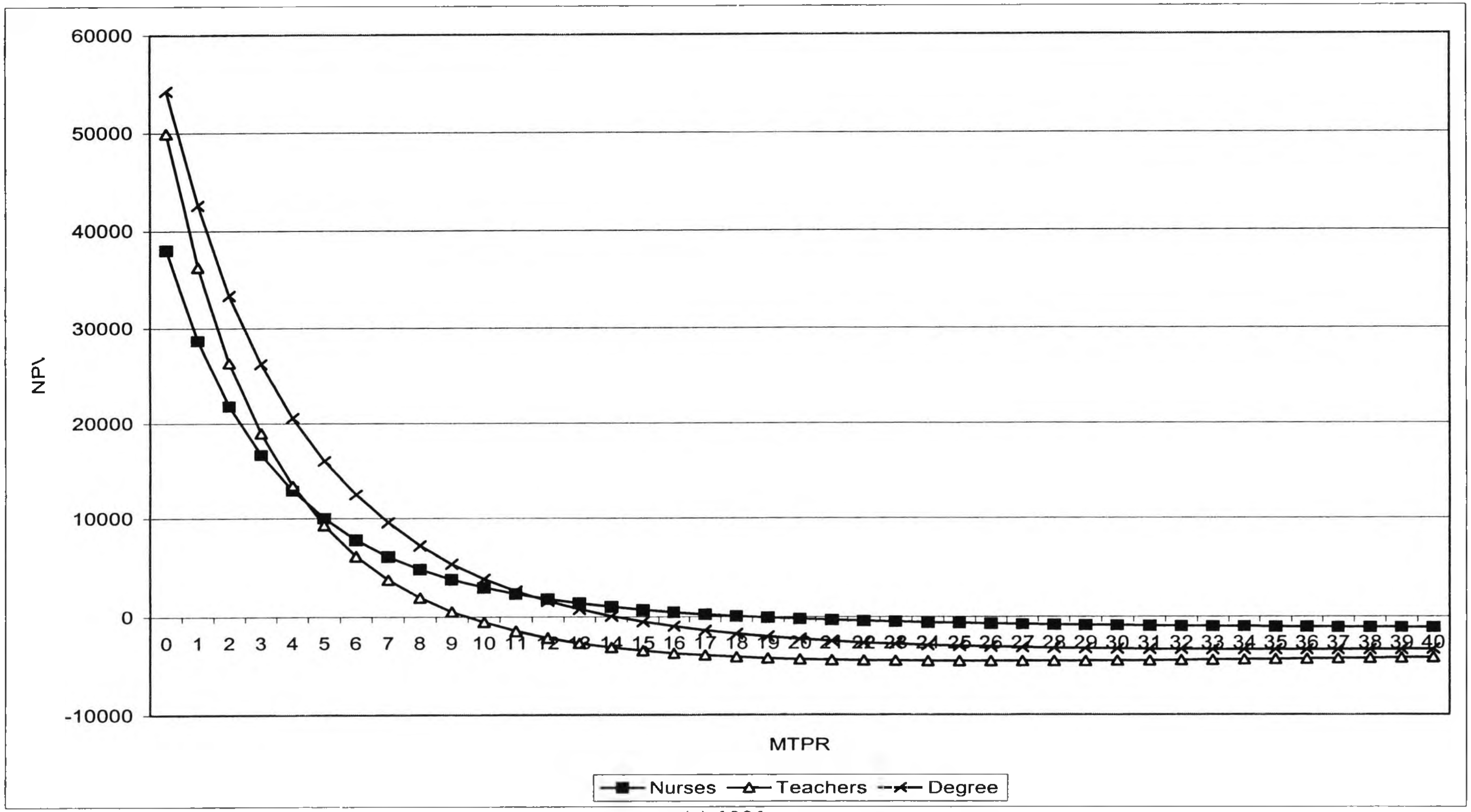
MTPR	Opportunity cost earnings = Non-manual workers			Opportunity cost earnings = All workers			Opportunity cost earnings = workers whose highest qualification is A level		
	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates
0	15,601	66,232	60,278	32,438	83,069	77,115	23,401	74,032	68,078
1	14,263	49,255	47,133	27,450	62,442	60,320	19,758	54,750	52,628
2	12,813	36,741	36,938	23,274	47,202	47,399	16,688	40,616	40,813
3	11,398	27,428	28,983	19,802	35,832	37,387	14,129	30,159	31,713
4	10,091	20,434	22,737	16,925	27,268	29,571	12,009	22,352	24,655
5	8,916	15,132	17,804	14,540	20,755	23,428	10,254	16,470	19,142
6	7,880	11,076	13,885	12,560	15,756	18,565	8,801	11,997	14,806
7	6,973	7,947	10,753	10,909	11,883	14,690	7,592	8,566	11,373
8	6,183	5,512	8,237	9,527	8,856	11,581	6,582	5,911	8,636
9	5,495	3,602	6,204	8,363	6,470	9,072	5,732	3,839	6,441
10	4,896	2,092	4,552	7,378	4,574	7,034	5,013	2,209	4,669
11	4,373	891	3,204	6,537	3,056	5,369	4,398	916	3,230
12	3,914	-71	2,099	5,816	1,831	4,001	3,869	-116	2,054
13	3,510	-846	1,189	5,194	838	2,872	3,410	-946	1,089
14	3,153	-1,474	436	4,653	26	1,935	3,009	-1,618	292
15	2,835	-1,985	-189	4,179	-641	1,155	2,655	-2,165	-370
16	2,552	-2,401	-710	3,763	-1,191	501	2,341	-2,612	-920
17	2,298	-2,742	-1,144	3,394	-1,646	-48	2,061	-2,979	-1,381
18	2,069	-3,022	-1,508	3,066	-2,025	-511	1,809	-3,282	-1,768
19	1,862	-3,251	-1,814	2,773	-2,340	-903	1,582	-3,532	-2,094
20	1,675	-3,439	-2,070	2,510	-2,604	-1,235	1,376	-3,739	-2,369
21	1,504	-3,593	-2,286	2,273	-2,824	-1,517	1,188	-3,910	-2,603
22	1,349	-3,718	-2,467	2,058	-3,008	-1,757	1,016	-4,051	-2,800
23	1,206	-3,820	-2,620	1,864	-3,162	-1,962	858	-4,168	-2,968
24	1,075	-3,902	-2,748	1,686	-3,291	-2,137	713	-4,264	-3,110
25	955	-3,966	-2,856	1,524	-3,397	-2,287	579	-4,342	-3,231
26	844	-4,017	-2,945	1,375	-3,486	-2,414	455	-4,406	-3,334
27	741	-4,055	-3,020	1,238	-3,558	-2,523	340	-4,456	-3,421
28	647	-4,083	-3,082	1,113	-3,617	-2,615	234	-4,496	-3,494
29	559	-4,102	-3,132	997	-3,664	-2,694	135	-4,526	-3,556
30	477	-4,114	-3,173	890	-3,701	-2,761	44	-4,548	-3,607
31	402	-4,119	-3,206	791	-3,730	-2,817	-42	-4,563	-3,649
32	332	-4,118	-3,231	699	-3,751	-2,864	-121	-4,571	-3,684
33	266	-4,113	-3,250	614	-3,765	-2,903	-195	-4,575	-3,712
34	205	-4,104	-3,264	535	-3,774	-2,935	-265	-4,573	-3,734
35	149	-4,091	-3,273	461	-3,778	-2,961	-329	-4,568	-3,751
36	96	-4,074	-3,278	393	-3,777	-2,981	-389	-4,559	-3,763
37	46	-4,056	-3,280	329	-3,773	-2,997	-445	-4,547	-3,771
38	0	-4,035	-3,278	270	-3,765	-3,009	-498	-4,532	-3,776
39	-43	-4,012	-3,274	214	-3,755	-3,017	-547	-4,515	-3,777
40	-83	-3,987	-3,267	162	-3,742	-3,022	-592	-4,496	-3,776

(e) 1995

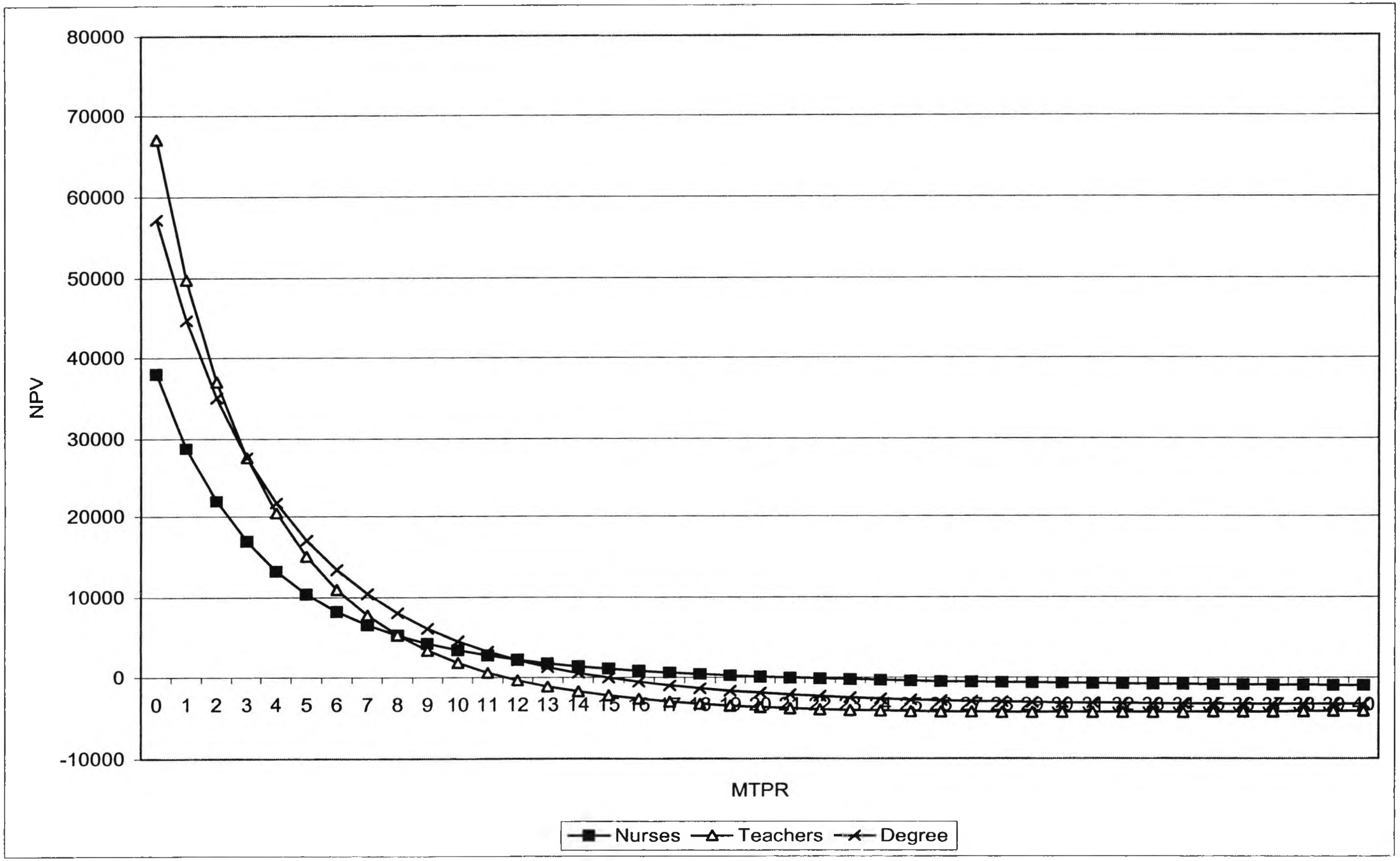
MTPR	Opportunity cost earnings = Non-manual workers			Opportunity cost earnings = All workers			Opportunity cost earnings = workers whose highest qualification is A level		
	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates	Nurses	Teachers	Graduates
0	34,967	64,769	62,061	53,196	82,999	80,290	61,292	91,094	88,386
1	27,306	47,768	47,684	41,622	62,084	62,000	48,347	68,808	68,725
2	21,666	35,271	36,676	33,058	46,662	48,068	38,718	52,323	53,728
3	17,459	25,999	28,187	26,642	35,182	37,370	31,468	40,007	42,195
4	14,277	19,055	21,595	21,774	26,552	29,091	25,937	30,715	33,254
5	11,835	13,808	16,441	18,030	20,003	22,635	21,661	23,634	26,267
6	9,934	9,808	12,385	15,113	14,987	17,564	18,312	18,187	20,763
7	8,431	6,733	9,174	12,809	11,110	13,551	15,653	13,955	16,396
8	7,226	4,348	6,617	10,965	8,087	10,356	13,514	10,636	12,905
9	6,245	2,485	4,568	9,470	5,710	7,794	11,770	8,010	10,094
10	5,435	1,019	2,920	8,242	3,827	5,727	10,330	5,914	7,815
11	4,758	-142	1,586	7,223	2,323	4,051	9,127	4,227	5,956
12	4,184	-1,068	503	6,366	1,113	2,684	8,110	2,858	4,429
13	3,693	-1,809	-380	5,638	135	1,565	7,242	1,739	3,168
14	3,269	-2,406	-1,102	5,014	-661	643	6,492	818	2,122
15	2,898	-2,887	-1,694	4,473	-1,312	-119	5,840	55	1,248
16	2,571	-3,277	-2,181	4,002	-1,846	-750	5,268	-580	516
17	2,282	-3,593	-2,581	3,588	-2,286	-1,275	4,762	-1,113	-101
18	2,024	-3,849	-2,911	3,222	-2,650	-1,713	4,312	-1,560	-623
19	1,792	-4,056	-3,182	2,896	-2,952	-2,078	3,909	-1,939	-1,065
20	1,583	-4,223	-3,405	2,604	-3,202	-2,384	3,546	-2,259	-1,442
21	1,393	-4,357	-3,588	2,342	-3,408	-2,640	3,218	-2,532	-1,763
22	1,221	-4,464	-3,737	2,105	-3,580	-2,853	2,921	-2,764	-2,037
23	1,064	-4,547	-3,858	1,891	-3,721	-3,032	2,650	-2,962	-2,273
24	921	-4,612	-3,956	1,696	-3,837	-3,181	2,403	-3,130	-2,474
25	789	-4,661	-4,033	1,518	-3,932	-3,305	2,176	-3,274	-2,647
26	668	-4,696	-4,094	1,355	-4,009	-3,407	1,967	-3,397	-2,795
27	557	-4,719	-4,140	1,206	-4,070	-3,491	1,775	-3,501	-2,923
28	455	-4,733	-4,175	1,070	-4,118	-3,560	1,598	-3,590	-3,031
29	360	-4,739	-4,199	944	-4,155	-3,615	1,434	-3,665	-3,125
30	273	-4,737	-4,214	828	-4,183	-3,659	1,283	-3,727	-3,204
31	192	-4,730	-4,221	720	-4,201	-3,693	1,142	-3,780	-3,271
32	117	-4,717	-4,223	621	-4,213	-3,718	1,012	-3,823	-3,328
33	47	-4,700	-4,218	529	-4,218	-3,736	890	-3,858	-3,375
34	-17	-4,680	-4,209	444	-4,218	-3,748	777	-3,885	-3,415
35	-77	-4,656	-4,196	365	-4,214	-3,754	672	-3,907	-3,447
36	-132	-4,629	-4,180	292	-4,205	-3,755	574	-3,923	-3,473
37	-184	-4,601	-4,161	224	-4,192	-3,753	483	-3,934	-3,494
38	-232	-4,570	-4,139	161	-4,177	-3,746	397	-3,940	-3,510
39	-277	-4,537	-4,115	101	-4,159	-3,737	318	-3,943	-3,521
40	-318	-4,504	-4,090	46	-4,139	-3,725	243	-3,942	-3,528

(f) 1996

Appendix 3.5. NPV of nurse training, teacher training and obtaining a degree with non-manual workers as opportunity cost earnings by MTPR in 1996

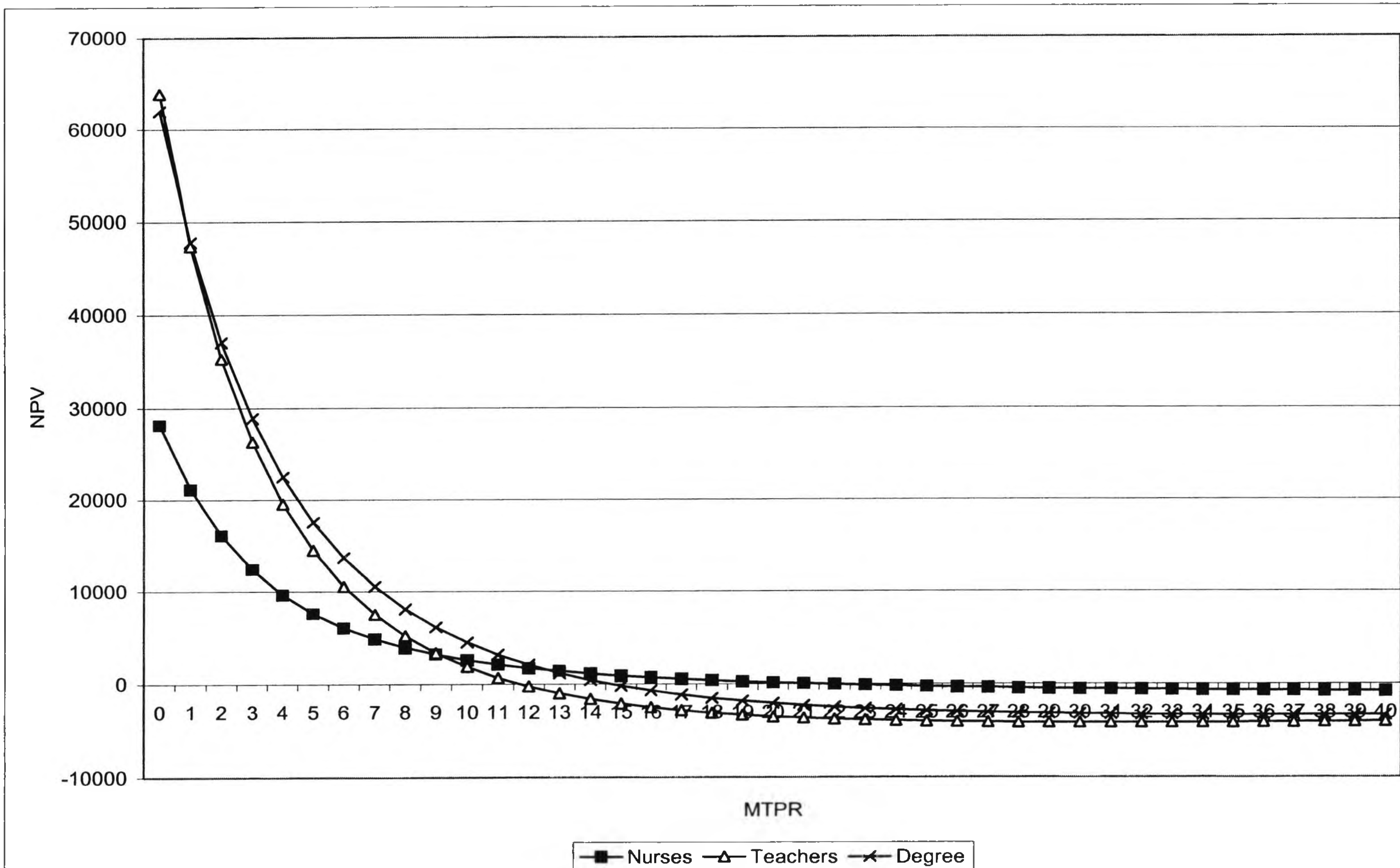


(a) 1991

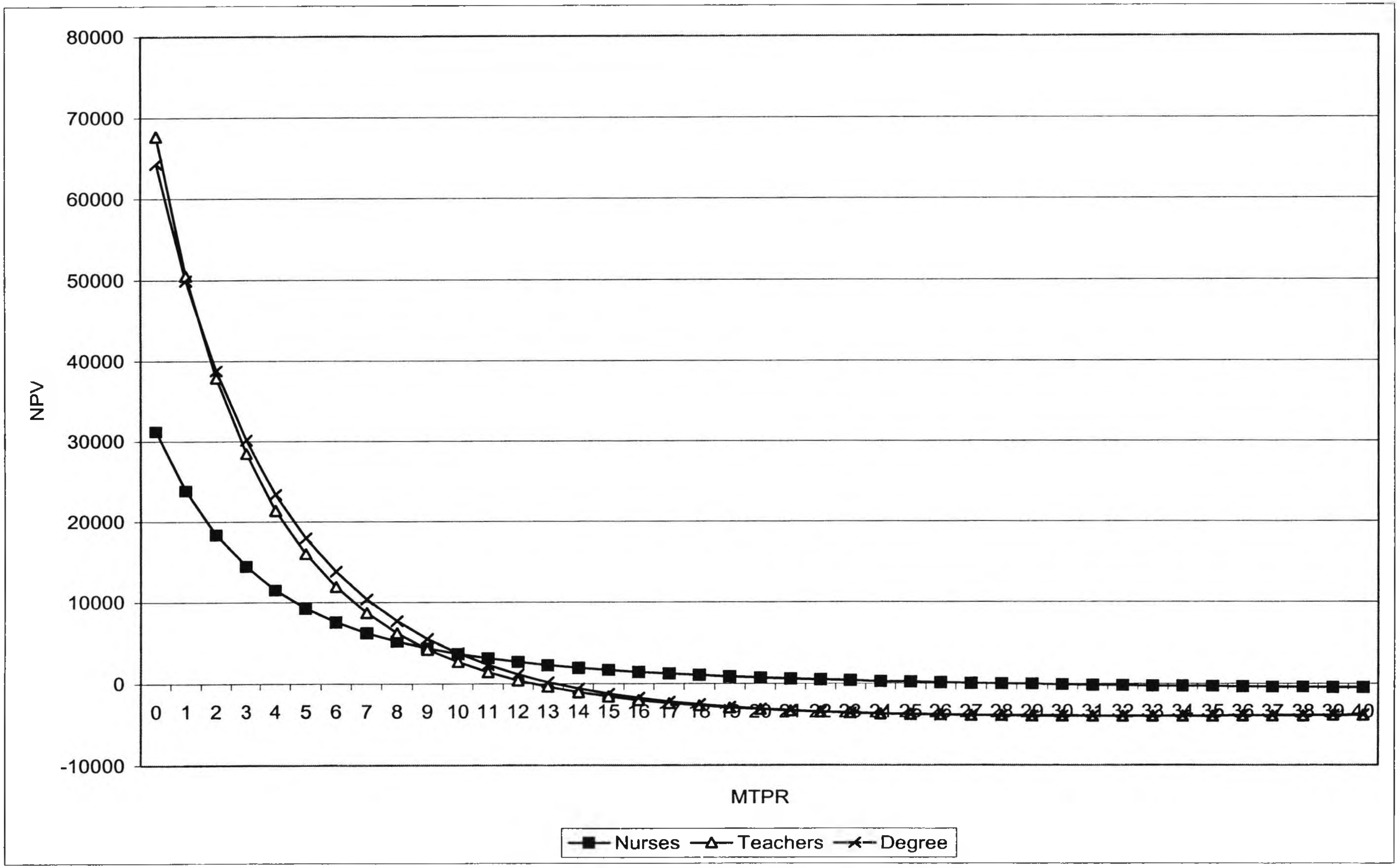


80000
70000
60000
50000
40000
30000
20000
10000
0
-10000
NPV
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
MTPR
Nurses
Teachers
Degree

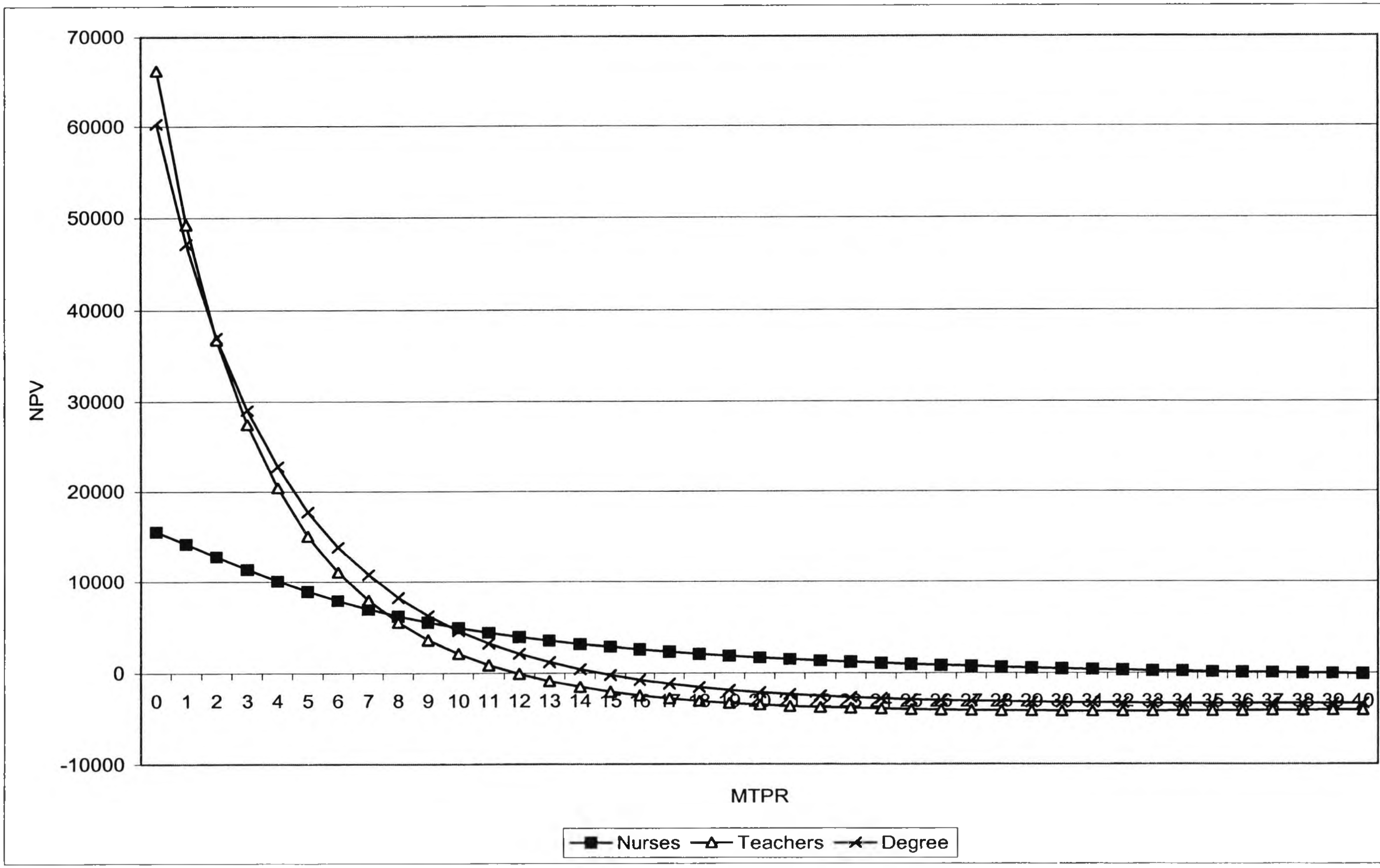
(b) 1992


70000
60000
50000
40000
30000
20000
10000
0
-10000
NPV
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
MTPR
Nurses
Teachers
Degree

(c) 1993


80000
70000
60000
50000
40000
30000
20000
10000
0
-10000
NPV
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
MTPR
Nurses
Teachers
Degree

(d) 1994


70000
60000
50000
40000
30000
20000
10000
0
-10000
NPV
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
MTPR
Nurses
Teachers
Degree

(e) 1995

APPENDICES TO CHAPTER 4

Appendix 4.1. Derivation of the Mincerian earnings function

The theoretical background for wage determination is human capital theory, which predicts that individual pay differences are the outcome of labour productivity differences arising from differences in human capital accumulation (Becker, 1964; Mincer, 1974). We wish to construct an economic model of wage determination in order to construct an earnings function for nurses and other workers.

Let C_t denote net investment in human capital in time period t, E_t denote potential earnings and Y_t denote actual observed earnings. Earnings in period 1 can be expressed in terms of prior investment in human capital. Suppose there is an investment in human capital of C_0 in period 0 (arising through, for example, a period of compulsory schooling). Potential earnings in the following year would be augmented by the returns r on the initial investment, so that

$$E_1 = E_0 + rC_0 \qquad [A4.1.1]$$

Similarly potential earnings in period 2 would equal earnings in period 1 plus the returns on investment. Thus:

$$E_2 = E_1 + rC_1 = E_0 + rC_0 + rC_1 \qquad [A4.1.2]$$

In general we have:

$$E_t = E_0 + r\sum_{i=0}^{t-1} C_i \qquad [A4.1.3]$$

In any time period, the size of the financial investment in human capital is equal to potential earnings minus observed earnings:

$$C_t = E_t - Y_t \qquad [A4.1.4]$$

Let s_t represent the proportion of time an individual spends investing in human capital. This can be expressed as:

$$s_t = C_t / E_t \qquad [A4.1.5]$$

which is the fraction of potential earnings E_t that the individual foregoes to accumulate human capital.

Substituting s_t for C_t yields:

$$E_1 = E_0 + rs_0E_0 = E_0(1 + rs_0) \qquad [A4.1.6]$$

and

$$E_2 = E_1 + rs_1E_1 = E_1(1+rs_1) = E_0(1 + rs_0)(1+rs_1) \qquad [A4.1.7]$$

More generally we have:

$$E_t = E_0(1 + rs_0)(1+rs_1)\ldots(1+rs_{t-1}) = \prod_{i=0}^{t-1}(1+rs_i) \qquad [A4.1.8]$$

Taking natural logarithms of both sides we have:

$$\ln E_t = \ln E_0 + \sum_{i=0}^{t-1}\ln(1+rs_i) \qquad [A4.1.9]$$

However, it is the case that $\ln(1 + x) \approx x$ when x is small. So, the above equation may be rewritten as:

$$\ln E_t = \ln E_0 + r_s\sum_{i=0}^{t-1}s_i \qquad [A4.1.10]$$

As described by equation [A4.1.5] the term s_t represents the proportion of time in each period spent investing in human capital. During the period of compulsory schooling s_t equals one since schooling is a full-time task. After schooling ends s_t declines. Thus s_t can be divided into two periods: a schooling period; and, a post-schooling period. It follows that equation [A4.1.10] can be decomposed into:

$$\ln E_t = \ln E_0 + r_s\sum_{i=0}^{S}s_i + r_p\sum_{i=S+1}^{t-1}s_i \qquad [A4.1.11]$$

where S represent years of schooling, r_s is the rate of return to schooling and r_p is the rate of return to post-school investments in human capital.

We can simplify this equation since $s_t = 1$ during the schooling period. This means that $\sum_{i=0}^{S} s_i = S$ and therefore:

$$\ln E_t = \ln E_0 + r_s S + r_p \sum_{i=S+1}^{t-1} s_i \qquad [A4.1.12]$$

where i now ranges over the years between leaving school and retirement.

It is the case that $s_t = 0$ when E_t is maximised. We can therefore construct the following relationship:

$$s_t = \alpha - \alpha t/m \qquad [A4.1.13]$$

where t now represents years of post-school work experience, α is the fraction of earnings capacity devoted to self-investment immediately on leaving school, and m is the number of years after leaving school at which potential earnings are maximised. Note that when potential earnings are maximised it is the case that m = t.

Substituting equation [A4.1.13] into equation [A4.1.12] we have:

$$\ln E_t = \ln E_0 + r_s S + r_p \sum_{i=S+1}^{t-1} (\alpha - \alpha i / m) \qquad [A4.1.14]$$

The third component of the right hand side of equation [A4.1.14] (an arithmetic progression) is approximated by:

$$\alpha t - [\alpha t(t-1)] / 2m \approx \alpha t - \alpha t^2 / 2m \qquad [A4.1.15]$$

Inserting this into equation [A4.1.14] we have:

$$\ln E_t = \ln E_0 + r_s S + r_p \alpha t - r_p \alpha t^2 / 2m \qquad [A4.1.16]$$

Now, we are unable to observe potential earnings E_t directly, but we know they are related to observed earnings Y_t in the following manner:

$$Y_t = E_t - C_{gt} \qquad [A4.1.17]$$

where C_{gt} is gross investment in human capital.

We have already specified the following relationship between s, C and E from equation [A4.1.5]:

$$s_{gt} = C_{gt} / E_t \qquad [A4.1.18]$$

where s_{gt} here is the time equivalent investment pertaining to gross investment in human capital. Thus:

$$Y_t = E_t(1 - s_{gt}) \qquad [A4.1.19]$$

It is also the case that:

$$C_t = C_{gt} - \delta PK_t \qquad [A4.1.20]$$

where δ is the rate of depreciation of human capital, and PK_t is the monetary value of human capital stock at time t (K_t is the accumulated human capital stock at time t, and P is the capitalised monetary value of one unit of human capital).

Combining equation [A4.1.18] with equation [A4.1.20] we have:

$$s_{gt} = (C_t + \delta PK_t) / E_t \qquad [A4.1.21]$$

Thus:

$$s_{gt} = C_t / E_t + \delta/r \qquad [A4.1.22]$$

since $E_t = rPK_t$, and therefore we have:

$$s_{gt} = s_t + \delta/r \qquad [A4.1.23]$$

Substituting this into equation [A4.1.8] we have:

$$Y_t = E_t(1 - s_t - \delta/r) \qquad [A4.1.24]$$

Taking natural logarithms of both sides:

$$\ln Y_t = \ln[E_t(1 - s_t - \delta/r)] \qquad [A4.1.25]$$

and since $\ln(1 + x) \approx x$ when x is small we have:

$$\ln Y_t = \ln E_t - s_t - \delta/r \qquad [A4.1.26]$$

Using equation [A1.13] we have:

$$\ln Y_t = \ln E_t - \alpha + \alpha t/m - \delta/r \qquad [A4.1.27]$$

Substituting this into equation [A4.1.16] gives the following:

$$\ln Y_t = [\ln E_0 - \alpha - \delta/r] + r_s S + [\alpha/m + r_p\alpha]t - [r_p\alpha / 2m]\, t^2 \qquad [A4.1.28]$$

Renaming the elements of equation [A4.1.28] gives the following, which is the same as equation [4.1] in the text:

$$\ln Y_t = \beta_0 + \beta_1 S + \beta_2 t + \beta_3 t^2 \qquad [A4.1.29]$$

where

$\beta_0 = \ln E_0 - \alpha - \delta/r$

$\beta_1 = r_s$

$\beta_2 = \alpha/m + r_p\alpha$

$\beta_3 = -[r_p\alpha / 2m]$

Appendix 4.2. Probit model of participation

The sample rule outlined in the text in equation [4.27]) is that $\ln W_i$ is observed only when P_i^* $= \delta Z_i + V_i$ is greater than zero (i.e. the offered wage is greater than the reservation wage). In other words;

$$H > 0 \text{ if and only if } V_i > -\delta Z_i \qquad [A4.2.1]$$

$$H = 0 \text{ if and only if } V_i \leq -\delta Z_i \qquad [A4.2.2]$$

This threshold condition means that i will work if and only if his wage exceeds his reservation wage; it also means that, given δZ_i, the value of V_i - individual i's taste for work - determines whether the individual works.

To proceed further, it is necessary to make an assumption of some kind about the distribution of tastes for work, V_i, in the population. It is convenient and reasonable to assume that V has a mean of zero and is normally distributed in the population as a whole. If so, then the transformed variable V/σ_V has a mean of zero and follows the standard normal distribution, where σ_V is the standard deviation of V_i in the population. This is shown in Figure A4.2.1, which plots the probability density $\phi(V/\sigma_V)$ of the standardized variable V_i/σ_V for each value of V_i/σ_V. Since the mean of V_i is zero (by assumption) the curve is centered at zero.

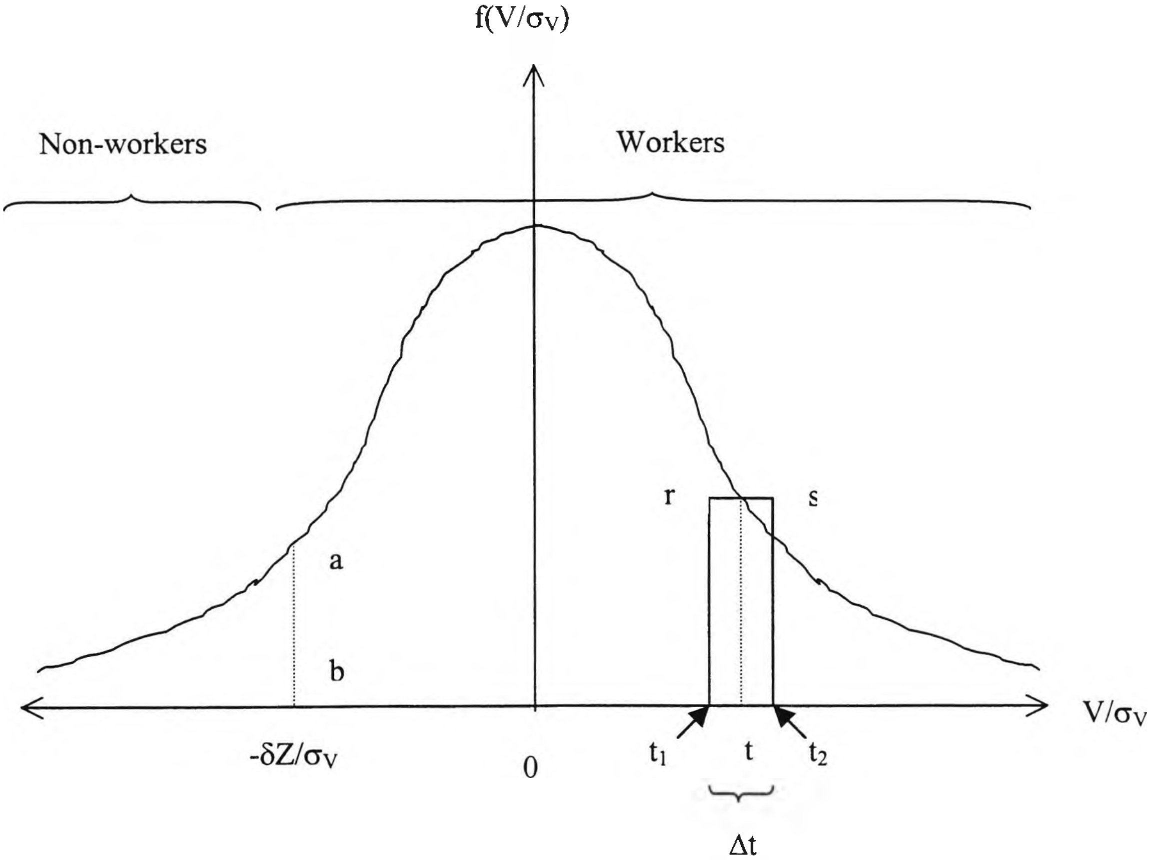


Figure A4.2.1. Illustration of probit analysis of participation

It is not possible to ascertain directly whether an individual with particular values of δZ will or will not work. That depends on V as well as on δZ and V, unlike δZ, is unobservable. However, given an assumption of normality about how V is distributed in the population, it is possible to derive an expression for the probability that such an individual will or will not work. In particular, the height of the curve in Figure A4.2.1 at any value of V_i/σ_V such as t is equal to the probability density of V_i/σ_V at that value, namely $\phi(t)$.

Consequently, the probability that a given individual i will participate (i.e. that $\ln W_i$ is observed) is given by:

$$P(\text{i participates}) = P(V_i/\sigma_V > -\delta Z/\sigma_V]) = \int_{-\delta Z_i/\sigma_V}^{\infty} \phi(t)dt = 1 - \Phi(-\delta Z/\sigma_V) \qquad [A4.2.3]$$

where ϕ is standard normal density function, and Φ in the standard normal cumulative density function.

$\phi(t)$ is (approximately) equal to the ratio of the probability that V_i/σ_V will lie within the upper and lower limits t_2 and t_1 of a small interval that includes the value t to the size of that small interval $t_2 - t_1$. Let Δt denote the quantity $t_2 - t_1$. Then:

$$\phi(t) = \lim_{t_2 - t_1 \to 0} \frac{P[t_2 > (V_i/\sigma_V) > t_1]}{t_2 - t_1} = \lim_{\Delta t \to 0} \frac{P[t_2 > (V_i/\sigma_V) > t_1]}{\Delta t} \qquad [A4.2.4]$$

So, for example, the probability that the value of V_i/σ_V will lie somewhere in between t_1 and t_2 in Figure A4.2.1 is approximated by the product of the height of the curve in the vicinity of point t [namely $\phi(t)$], and the distance $\Delta t = t_2 - t_1$; that is, by the area $t_1\, t_2rs$. Therefore:

$$P[t_2 > V_i/\sigma_V > t_1\,] \approx \phi(t)\Delta t \qquad [A4.2.5]$$

By extension, the probability that V_i/σ_V will exceed $-\delta Z/\sigma_V$ is approximated by adding the areas of the many rectangles similar to $t_1\, t_2rs$ that one could construct under the curve and to the right of ab. This is simply the entire area under the curve and to the right of ab, described by equation [A4.2.3].

In sum, the probability that someone will work, when his value of $-\delta Z/\sigma_V$ is equal to the amount associated with the line ab, is given by the area under the curve in Figure A4.2.1 to the right of ab. By extension, suppose one were to consider a large number of individuals all of whom had a value of $-\delta Z/\sigma_V$ equal to the amount associated with the line ab. The area under the curve to the right of ab gives the proportion of these individuals who would work.

Note that when δZ is negative and large in absolute value, the threshold value $-\delta Z/\sigma_V$ will be positive and large, the threshold line ab will lie to the right of the centre of the normal distribution curve in Figure A4.2.1, and only a small proportion of individuals with that value of δZ_i and above will work. Such individuals will all have very strong tastes for work, (i.e. they will have large positive values of V_i). This is directly related to the notion of sample selection bias: other things being equal, restricting empirical analysis to persons who work may result in a sample in which V_i does not have a mean of zero even if V_i does have a mean of zero in the total population.

The assumption that V_i is a normally distributed random variable leads directly to an participation equation suitable for empirical estimation, whose parameters may be estimated by maximum likelihood methods. The likelihood functions for a sample of persons who are either employed or not employed are given as:

$$l = \prod_{i \in E} [1 - \Phi(-\delta Z / \sigma_V)] * \prod_{i \in \bar{E}} \Phi(-\delta Z / \sigma_V) \qquad \text{[A4.2.6]}$$

where E is the set of persons who participate, and $\bar{E}$ is the set of persons who do not participate.

This is the probit equation for participation in the labour force, whose parameters δ may be estimated by maximizing the likelihood function (or its logarithm) presented in equation [A4.2.6] with respect to δ.

APPENDICES TO CHAPTER 5

Appendix 5.1. Time period to which each of the quarters used in the statistical models pertain

Quarter	Time period to which quarter pertains [1]
0	Winter 1992/3
1	Spring 1993
2	Summer 1993
3	Autumn 1993
4	Winter 1993/4
5	Spring 1994
6	Summer 1994
7	Autumn 1994
8	Winter 1994/5
9	Spring 1995
10	Summer 1995
11	Autumn 1995
12	Winter 1995/6
13	Spring 1996
14	Summer 1996
15	Autumn 1996
16	Winter 1996/7
17	Spring 1997
18	Summer 1997
19	Autumn 1997
20	Winter 1997/8
21	Spring 1998
22	Summer 1998
23	Autumn 1998
24	Winter 1998/9
25	Spring 1999
26	Summer 1999
27	Autumn 1999
28	Winter 1999/2000
29	Spring 2000
30	Summer 2000
31	Autumn 2000

[1] The quarters are defined as follows:
Spring = March to May
Summer = June to August
Autumn = September to November
Winter = December to February

Appendix 5.2. Descriptive statistics of variables used in the participation and occupation selection equations

The definitions and descriptive statistics of the variables used in the participation equations for Model 2 (Table A5.2.1) and Model 4 (Table A5.2.2), and the occupation selection equation for Model 5 (Table A5.2.3) are given below. Table A5.2.4 provides descriptive statistics of variables used in the occupation selection and participation equations across different sub-samples of the population. From Table A5.2.1 we note that the entire sample of all workers and non-workers used in the analysis consist of 247,774 females aged 18 to 60 years. Of these, 61% participate in the labour market. Note that across this period in Great Britain some 70% of all females participated in the labour market (ONS, 2000). Therefore, the QLFS sample used here contains a slightly lower proportion of participating females than the British population. However, as outlined above, this is explained by the exclusion of the self-employed from the dataset due to a lack of wage information in this sub-group of the participating population.

	All workers + non-workers		Definition
	Mean	Std.Dev.	
PART	0.6132	0.4870	Participate in the labour market=1, 0 otherwise
Age variables			
AGE	38.2208	11.5409	Years of age
AGE2	1594.0200	901.3600	Years of age squared
Personal characteristic variables			
DISABLE	0.1246	0.3302	Health problems affect paid work =1, 0 otherwise
ETHNIC	0.0577	0.2332	Non-white ethnic group=1, 0 otherwise
NONBRIT	0.0513	0.2206	Non-British nationality=1, 0 otherwise
ETHNBRIT	0.0193	0.1375	Non-white and non-British=1, 0 otherwise
Family variables			
PCHILD	0.1483	0.3554	Age 20-29 years and cohabiting or age 25-34 years and married=1, 0 otherwise
COHABIT	0.0896	0.2856	Cohabiting (living as a couple but not married)=1, 0 otherwise
MARRIED	0.5920	0.4915	Married=1, 0 otherwise
Property income variables			
PENSION	0.0159	0.1251	Receives an occupational pension=1, 0 otherwise
NONLABY	111.3340	951.3130	Non-labour income

Years of education variables			
YED	13.0982	2.7066	Years of full-time education
YED2	178.8880	80.2600	Years of full-time education squared
Educational attainment variables			
PGDEG	0.0231	0.1501	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG	0.0723	0.2589	Highest qualification is a first degree=1, 0 otherwise
ALEVEL	0.0707	0.2564	Highest qualification is A level=1, 0 otherwise
NOQUAL	0.2321	0.4221	Has no qualifications=1, 0 otherwise
Regional variables			
SEAST	0.2985	0.4576	Lives in the South East of England=1, 0 otherwise
Time trend variables			
Q1	0.0278	0.1644	Quarter 1=1, 0 otherwise
Q2	0.0277	0.1642	Quarter 2=1, 0 otherwise
Q3	0.0745	0.2625	Quarter 3=1, 0 otherwise
Q8	0.0743	0.2623	Quarter 8=1, 0 otherwise
Q13	0.0723	0.2590	Quarter 13=1, 0 otherwise
Q17	0.0280	0.1651	Quarter 17=1, 0 otherwise
Q18	0.0284	0.1661	Quarter 18=1, 0 otherwise
Q19	0.0277	0.1642	Quarter 19=1, 0 otherwise
Q20	0.0276	0.1637	Quarter 20=1, 0 otherwise
Q21	0.0273	0.1630	Quarter 21=1, 0 otherwise
Q22	0.0279	0.1647	Quarter 22=1, 0 otherwise
Q23	0.0276	0.1638	Quarter 23=1, 0 otherwise
Q24	0.0276	0.1639	Quarter 24=1, 0 otherwise
Q25	0.0268	0.1615	Quarter 25=1, 0 otherwise
Q26	0.0269	0.1619	Quarter 26=1, 0 otherwise
Q27	0.0267	0.1613	Quarter 27=1, 0 otherwise
Q28	0.0526	0.2232	Quarter 28=1, 0 otherwise
Q29	0.0514	0.2208	Quarter 29=1, 0 otherwise
Q30	0.0508	0.2195	Quarter 30=1, 0 otherwise
Q31	0.0502	0.2184	Quarter 31=1, 0 otherwise
N	247,774		

Table A5.2.1 Descriptive statistics of variables in participation equation for Model 2

In Table A5.2.2 we present descriptive statistics for variables used in the separate participation equations for nurses and all other workers in Model 4. Using the definition of non-participating nurses given above (individuals with a nursing qualification who do not work) we can see that the participation rate is higher for the 8,878 nurses in the data than for the 238,896 other workers (74% versus 61%).

	Nurses [1]		All Other Workers [1]		Definition
	Mean	Std.Dev.	Mean	Std.Dev.	
PART*	0.7443	0.4363	0.6084	0.4881	Participate in the labour market=1, 0 otherwise

Age variables					
AGE*	40.2184	10.0635	38.1465	11.5855	Years of age
AGE2*	1718.7800	832.8290	1589.3800	903.4760	Years of age squared
Personal characteristic variables					
DISABLE*	0.1107	0.3138	0.1251	0.3308	Health problems affect paid work =1, 0 otherwise
ETHNIC#	0.0534	0.2248	0.0579	0.2335	Non-white ethnic group=1, 0 otherwise
NONBRIT*	0.0596	0.2367	0.0510	0.2199	Non-British nationality=1, 0 otherwise
ETHNBRIT	0.0182	0.1339	0.0193	0.1376	Non-white and non-British=1, 0 otherwise
Family variables					
PCHILD*	0.1875	0.3904	0.1468	0.3539	Age 20-29 years and cohabiting or age 25-34 years and married=1, 0 otherwise
COHABIT*	0.0714	0.2575	0.0903	0.2865	Cohabiting (living as a couple but not married)=1, 0 otherwise
MARRIED*	0.6791	0.4669	0.5888	0.4921	Married=1, 0 otherwise
Property income variables					
PENSION*	0.0286	0.1667	0.0154	0.1233	Receives an occupational pension=1, 0 otherwise
NONLABY	116.2700	788.3190	111.1510	956.8360	Non-labour income
Years of education variables					
YED*	13.4759	2.0353	13.0841	2.7274	Years of full-time education
YED2*	185.7420	62.9678	178.6330	80.8202	Years of full-time education squared
Educational attainment variables					
PGDEG*	0.0124	0.1106	0.0235	0.1513	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG*	0.0601	0.2378	0.0727	0.2597	Highest qualification is a first degree=1, 0 otherwise
ALEVEL*	2.0000	2.0000	0.0731	0.2603	Highest qualification is A level=1, 0 otherwise
NOQUAL*	2.0000	2.0000	0.2406	0.4274	Has no qualifications=1, 0 otherwise
Regional variables					
SEAST*	0.2742	0.4461	0.2994	0.4580	Lives in the South East of England=1, 0 otherwise
Time trend variables					
Q1	0.0273	0.1628	0.0278	0.1644	Quarter 1=1, 0 otherwise
Q2	0.0292	0.1683	0.0277	0.1641	Quarter 2=1, 0 otherwise
Q3*	0.0623	0.2417	0.0749	0.2633	Quarter 3=1, 0 otherwise
Q8*	0.0595	0.2365	0.0749	0.2632	Quarter 8=1, 0 otherwise
Q13*	0.0590	0.2357	0.0728	0.2598	Quarter 13=1, 0 otherwise
Q17	0.0300	0.1705	0.0280	0.1649	Quarter 17=1, 0 otherwise
Q18	0.0310	0.1733	0.0283	0.1658	Quarter 18=1, 0 otherwise
Q19	0.0270	0.1622	0.0278	0.1643	Quarter 19=1, 0 otherwise
Q20	0.0258	0.1585	0.0276	0.1639	Quarter 20=1, 0 otherwise
Q21	0.0265	0.1605	0.0273	0.1631	Quarter 21=1, 0 otherwise
Q22	0.0306	0.1723	0.0278	0.1644	Quarter 22=1, 0 otherwise
Q23	0.0266	0.1609	0.0276	0.1639	Quarter 23=1, 0 otherwise
Q24	0.0256	0.1579	0.0277	0.1642	Quarter 24=1, 0 otherwise
Q25	0.0271	0.1625	0.0268	0.1615	Quarter 25=1, 0 otherwise
Q26	0.0260	0.1592	0.0270	0.1620	Quarter 26=1, 0 otherwise
Q27#	0.0297	0.1699	0.0266	0.1609	Quarter 27=1, 0 otherwise
Q28#	0.0489	0.2156	0.0527	0.2235	Quarter 28=1, 0 otherwise
Q29	0.0513	0.2205	0.0514	0.2208	Quarter 29=1, 0 otherwise
Q30	0.0490	0.2159	0.0508	0.2196	Quarter 30=1, 0 otherwise
Q31	0.0531	0.2242	0.0501	0.2182	Quarter 31=1, 0 otherwise
N	8,878		238,896		

[1] Includes workers and non-workers

* Difference in mean values between nurses and all other workers significant at the 5% level

Difference in mean values between nurses and all other workers significant at the 10% level

Table A5.2.2. Descriptive statistics of variables used in participation equations for Model 4

Table A5.2.3 presents the descriptive statistics of variables used in the occupation selection equation in Model 5. Across all workers in the data (151,944 individuals) 4% are employed as nurses. 3% of all 247,774 individuals (workers and non-workers) are employed as nurses. This is equivalent to other estimates of the proportion of females who are employed as nurses in Great Britain (OHE, 2000), and implies that the data is representative of the general population in this regard.

	All workers only		Description
	Mean	Std.Dev.	
NURSE	0.0435	0.2040	Employed as a nurse=1, 0 otherwise
NURSEQUA	0.0633	0.2435	Has a nursing qualification=1, 0 otherwise
Age variables			
AGE	38.1904	10.9732	Years of age
AGE2	1578.9200	852.1770	Years of age squared
Personal characteristic variables			
DISABLE	0.0525	0.2230	Health problems affect paid work =1, 0 otherwise
ETHNIC	0.0359	0.1860	Non-white ethnic group=1, 0 otherwise
NONBRIT	0.0366	0.1878	Non-British nationality=1, 0 otherwise
ETHNBRIT	0.0093	0.0958	Non-white and non-British=1, 0 otherwise
Family variables			
PCHILD	0.1504	0.3575	Age 20-29 years and cohabiting or age 25-34 years and married=1, 0 otherwise
COHABIT	0.1033	0.3044	Cohabiting (living as a couple but not married)=1, 0 otherwise
MARRIED	0.6099	0.4878	Married=1, 0 otherwise
Property income variables			
PENSION	0.0116	0.1070	Receives an occupational pension=1, 0 otherwise
NONLABY	115.5410	964.1210	Non-labour income
Years of education variables			
YED	13.2908	2.5012	Years of full-time education
YED2	182.9020	75.1721	Years of full-time education squared
Educational attainment variables			
PGDEG	0.0315	0.1747	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG	0.0922	0.2894	Highest qualification is a first degree=1, 0 otherwise
ALEVEL	0.0685	0.2526	Highest qualification is A level=1, 0 otherwise
NOQUAL	0.1540	0.3609	Has no qualifications=1, 0 otherwise
Regional variables			
SEAST	0.3031	0.4596	Lives in the South East of England=1, 0 otherwise
Time trend variables			
Q1	0.0270	0.1622	Quarter 1=1, 0 otherwise

Q2	0.0261	0.1595	Quarter 2=1, 0 otherwise
Q3	0.0265	0.1605	Quarter 3=1, 0 otherwise
Q4	0.0262	0.1597	Quarter 4=1, 0 otherwise
Q5	0.0266	0.1610	Quarter 5=1, 0 otherwise
Q6	0.0277	0.1640	Quarter 6=1, 0 otherwise
Q7	0.0277	0.1640	Quarter 7=1, 0 otherwise
Q8	0.0280	0.1651	Quarter 8=1, 0 otherwise
Q9	0.0269	0.1617	Quarter 9=1, 0 otherwise
Q10	0.0283	0.1658	Quarter 10=1, 0 otherwise
Q11	0.0286	0.1666	Quarter 11=1, 0 otherwise
Q12	0.0289	0.1674	Quarter 12=1, 0 otherwise
Q13	0.0286	0.1665	Quarter 13=1, 0 otherwise
Q14	0.0284	0.1660	Quarter 14=1, 0 otherwise
Q15	0.0277	0.1641	Quarter 15=1, 0 otherwise
Q16	0.0289	0.1675	Quarter 16=1, 0 otherwise
Q17	0.0293	0.1688	Quarter 17=1, 0 otherwise
Q18	0.0296	0.1695	Quarter 18=1, 0 otherwise
Q19	0.0295	0.1691	Quarter 19=1, 0 otherwise
Q20	0.0287	0.1670	Quarter 20=1, 0 otherwise
Q21	0.0290	0.1678	Quarter 21=1, 0 otherwise
Q22	0.0297	0.1698	Quarter 22=1, 0 otherwise
Q23	0.0294	0.1689	Quarter 23=1, 0 otherwise
Q24	0.0295	0.1692	Quarter 24=1, 0 otherwise
Q25	0.0283	0.1658	Quarter 25=1, 0 otherwise
Q26	0.0288	0.1671	Quarter 26=1, 0 otherwise
Q27	0.0285	0.1664	Quarter 27=1, 0 otherwise
Q28	0.0536	0.2252	Quarter 28=1, 0 otherwise
Q29	0.0523	0.2227	Quarter 29=1, 0 otherwise
Q30	0.0533	0.2246	Quarter 30=1, 0 otherwise
Q31	0.0516	0.2212	Quarter 31=1, 0 otherwise
N	151,944		

Table A5.2.3. Descriptive statistics of variables in occupation selection equation for Model 5

Table A5.2.4 presents values for variables used in the participation and occupation selection equations for four sub-groups of the population (all workers, all non-workers, nurses only and all other workers only). As discussed in the main text, the key variables in the participation equations are the property income variables between workers and non-workers. In terms of differences in property income variables between workers and non-workers, as one might expect a (very slightly) greater proportion of non-workers receive an occupation pension (1% for workers, 2% for non-workers). The mean non-labour income of workers is in fact, on average, greater than that of non-workers (£115 per annum versus £104 per annum). This is perhaps surprising given that one might generally expect property income to be inversely related to participation. This might be explained by two factors. First, non-labour income measured here is an incomplete measure of property income (for example, it does not include the occupational pension figures, which are presented separately). Second, this might reflect the fact that the utility function of workers and non-workers with respect to the participation decision is different. As can be seen from Figure 4.1 in Chapter 4 an increase in property income N would generally lead to an increase in leisure L for a specific individual with a particular set of indifference curves. Further, the equilibrium amount of leisure L might be different for two different individuals for a given value of N – this will depend on the shape of their indifference curves. By extension, an individual with a high level of property income N may actually have a lower equilibrium amount of leisure L than an individual with a low level of property income if their indifference curves are suitably different.

The key variables in the occupation selection equations are the years of education and educational attainment variables between nurses and all other workers. These are important as measures of individual productive characteristics. Nurses and other workers have comparable years of full-time education (mean approximately 13 years). As one would expect, a much greater proportion of nurses have a nursing qualification than the rest of the working population (92% of nurses versus 2% of all other workers); this is clearly unsurprising given that having a nursing qualification is likely to be a basic requirement for being employed as a nurse. Indeed, what is perhaps surprising is the fact that some 8% of nurses do not have a nursing qualification at all. At the top end of the educational attainment spectrum nurses are less well educated than other workers in the sense that a lower proportion of nurses have postgraduate degrees (1% versus 3%), first degrees (6% versus 9%) and A levels (1% versus 7%) as their highest educational qualification. However, at the other end of the spectrum nurses are better educated than other workers because a greater proportion of nurses do possess some form of educational qualification (99% versus 84%).

	All workers		Non-workers		Nurses [1]		All other workers [1]	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
PART [2]	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000	1.0000	0.0000
NURSE [2]	0.0435	0.2040	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Age variables								
AGE*	38.1903	10.9731	38.2689	12.3877	38.8452	9.5285	38.1605	11.0334
AGE2*	1578.9100	852.1700	1617.9600	973.7920	1599.7300	767.8380	1577.9600	855.7970
Personal characteristic variables								
DISABLE#	0.0525	0.2230	0.2389	0.4264	0.0478	0.2134	0.0527	0.2234
ETHNIC*	0.0359	0.1860	0.0923	0.2894	0.0528	0.2237	0.0351	0.1841
NONBRIT*	0.0366	0.1878	0.0745	0.2626	0.0605	0.2385	0.0355	0.1851
ETHNBRIT*	0.0093	0.0958	0.0352	0.1842	0.0182	0.1335	0.0089	0.0937
Family variables								
PCHILD*	0.1504	0.3575	0.1448	0.3519	0.1984	0.3988	0.1483	0.3554
COHABIT*	0.1033	0.3044	0.0678	0.2514	0.0867	0.2814	0.1041	0.3054
MARRIED*	0.6099	0.4878	0.5636	0.4959	0.6571	0.4747	0.6078	0.4882
Property income variables								
PENSION	0.0116	0.1070	0.0227	0.1491	0.0118	0.1080	0.0116	0.1070
NONLABY	115.5410	964.1240	104.6640	930.6090	111.1930	774.2600	115.7390	971.8770
Years of education variables								
YED*	13.2908	2.5012	12.7928	2.9781	13.5051	1.9672	13.2811	2.5224
YED2*	182.9020	75.1722	172.5240	87.3475	186.2580	59.3591	182.7500	75.8094
Educational attainment								
NURSEQUA*	0.0633	0.2435	0.0237	0.1521	0.9215	0.2690	0.0243	0.1539
PGDEG*	0.0315	0.1747	0.0097	0.0979	0.0135	0.1153	0.0323	0.1768
DEG*	0.0922	0.2894	0.0406	0.1973	0.0642	0.2451	0.0935	0.2912
ALEVEL*	0.0685	0.2526	0.0743	0.2623	0.0101	0.1002	0.0712	0.2571
NOQUAL*	0.1540	0.3609	0.3559	0.4788	0.0045	0.0672	0.1608	0.3673
Regional variables								
SEAST*	0.3031	0.4596	0.2912	0.4543	0.2624	0.4400	0.3049	0.4604
Time trend variables								
Q1	0.0270	0.1622	0.0290	0.1677	0.0280	0.1650	0.0270	0.1621
Q2	0.0261	0.1595	0.0303	0.1715	0.0277	0.1641	0.0260	0.1592
Q3#	0.0265	0.1605	0.1506	0.3577	0.0304	0.1717	0.0263	0.1600
Q4	0.0262	0.1597	0.0000	0.0000	0.0259	0.1588	0.0262	0.1598
Q5	0.0266	0.1610	0.0000	0.0000	0.0286	0.1667	0.0266	0.1608
Q6	0.0277	0.1640	0.0000	0.0000	0.0306	0.1722	0.0275	0.1636
Q7	0.0277	0.1640	0.0000	0.0000	0.0269	0.1619	0.0277	0.1641
Q8	0.0280	0.1651	0.1477	0.3548	0.0294	0.1688	0.0280	0.1649
Q9#	0.0269	0.1617	0.0000	0.0000	0.0306	0.1722	0.0267	0.1613
Q10	0.0283	0.1658	0.0000	0.0000	0.0286	0.1667	0.0283	0.1658
Q11	0.0286	0.1666	0.0000	0.0000	0.0269	0.1619	0.0286	0.1668
Q12	0.0289	0.1674	0.0000	0.0000	0.0265	0.1606	0.0290	0.1677
Q13	0.0286	0.1665	0.1416	0.3487	0.0292	0.1684	0.0285	0.1665
Q14	0.0284	0.1660	0.0000	0.0000	0.0316	0.1750	0.0282	0.1656
Q15	0.0277	0.1641	0.0000	0.0000	0.0272	0.1628	0.0277	0.1642
Q16#	0.0289	0.1675	0.0000	0.0000	0.0254	0.1574	0.0291	0.1680
Q17	0.0293	0.1688	0.0260	0.1591	0.0295	0.1692	0.0293	0.1688

Q18	0.0296	0.1695	0.0265	0.1606	0.0328	0.1782	0.0294	0.1691
Q19	0.0295	0.1691	0.0250	0.1562	0.0286	0.1667	0.0295	0.1692
Q20	0.0287	0.1670	0.0258	0.1584	0.0281	0.1654	0.0287	0.1670
Q21	0.0290	0.1678	0.0246	0.1550	0.0262	0.1597	0.0291	0.1681
Q22	0.0297	0.1698	0.0250	0.1562	0.0315	0.1746	0.0296	0.1696
Q23	0.0294	0.1689	0.0247	0.1553	0.0280	0.1650	0.0295	0.1691
Q24	0.0295	0.1692	0.0247	0.1551	0.0272	0.1628	0.0296	0.1695
Q25	0.0283	0.1658	0.0245	0.1545	0.0268	0.1615	0.0284	0.1660
Q26	0.0288	0.1671	0.0241	0.1532	0.0269	0.1619	0.0288	0.1674
Q27	0.0285	0.1664	0.0239	0.1527	0.0315	0.1746	0.0284	0.1660
Q28*	0.0536	0.2252	0.0510	0.2200	0.0483	0.2144	0.0538	0.2257
Q29	0.0523	0.2227	0.0499	0.2177	0.0525	0.2231	0.0523	0.2227
Q30	0.0533	0.2246	0.0467	0.2111	0.0518	0.2216	0.0534	0.2248
Q31	0.0516	0.2212	0.0480	0.2138	0.0540	0.2261	0.0515	0.2210
N	151,944		95,830		6,608		145,335	

[1] Does not include non-workers
[2] Not included in comparison of means since used to define compared samples
* Difference in mean values between nurses and all other workers significant at the 5% level
[#] Difference in mean values between nurses and all other workers significant at the 10% level

Table A5.2.4. Descriptive statistics of variables in occupation selection and participation equations for sub-groups of the population

The full set of descriptive statistics for the variables included in the wage equations (including the time trend variables) are presented in Table A5.2.5.

	All workers [1]		Nurses only [2]		Other workers only [2]		Definition
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	
LNWAGE*	1.5830	0.5138	1.9320	0.3835	1.5671	0.5133	LN hourly wage
NURSE	0.0435	0.2040					Employed as a nurse=1, 0 otherwise
Years of education variables							
YED*	13.2908	2.5012	13.5051	1.9672	13.2811	2.5224	Years of full-time education
YED2*	182.9020	75.1721	186.2580	59.3591	182.7490	75.8093	Years of full-time education squared
Educational attainment variables							
NURSEQUA*			0.9215	0.2690	0.0243	0.1539	Has a nursing qualification=1, 0 otherwise
PGDEG*	0.0315	0.1747	0.0135	0.1153	0.0323	0.1768	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG*	0.0922	0.2894	0.0642	0.2451	0.0935	0.2912	Highest qualification is a first degree=1, 0 otherwise
ALEVEL*	0.0685	0.2526	0.0101	0.1002	0.0712	0.2571	Highest qualification is A level=1, 0 otherwise
NOQUAL*	0.1540	0.3609	0.0045	0.0672	0.1608	0.3673	Has no qualifications=1, 0 otherwise

Work experience variables							
EXP*	9.7689	7.0677	11.9953	7.9082	9.6677	7.0103	Years of experience with current employer
EXP2*	145.3830	210.2600	206.4170	253.3180	142.6080	207.6660	Years of experience with current employer squared
Personal characteristics variables							
DISABLE[#]	0.0525	0.2230	0.0478	0.2134	0.0527	0.2234	Health problems affect paid work =1, 0 otherwise
ETHNIC*	0.0359	0.1860	0.0528	0.2237	0.0351	0.1841	Non-white ethnic group=1, 0 otherwise
NONBRIT*	0.0366	0.1878	0.0605	0.2385	0.0355	0.1851	Non-British nationality=1, 0 otherwise
ETHNBRIT*	0.0093	0.0958	0.0182	0.1335	0.0089	0.0937	Non-white and non-British=1, 0 otherwise
Regional variables							
SEAST*	0.3031	0.4596	0.2624	0.4400	0.3049	0.4604	Lives in the South East of England=1, 0 otherwise
Job characteristic variables							
HOURSPW*	31.2637	12.9985	33.5796	10.5943	31.1584	13.0876	Total usual hours worked per week
MANAGE*	0.2696	0.4438	0.7639	0.4247	0.2472	0.4314	Employed as a supervisor, manager or foreman=1, 0 otherwise
NWORKERS*	0.6347	0.4815	0.8390	0.3676	0.6254	0.4840	25+ workers at workplace=1, 0 otherwise
TEMP*	0.0756	0.2643	0.0610	0.2393	0.0762	0.2654	Job is non-permanent or temporary=1, 0 otherwise
Time trend variables							
Q1	0.0270	0.1622	0.0280	0.1650	0.0270	0.1621	Quarter 1=1, 0 otherwise
Q2	0.0261	0.1595	0.0277	0.1641	0.0260	0.1592	Quarter 2=1, 0 otherwise
Q3[#]	0.0265	0.1605	0.0304	0.1717	0.0263	0.1600	Quarter 3=1, 0 otherwise
Q4	0.0262	0.1597	0.0259	0.1588	0.0262	0.1598	Quarter 4=1, 0 otherwise
Q5	0.0266	0.1610	0.0286	0.1667	0.0266	0.1608	Quarter 5=1, 0 otherwise
Q6	0.0277	0.1640	0.0306	0.1722	0.0275	0.1636	Quarter 6=1, 0 otherwise
Q7	0.0277	0.1640	0.0269	0.1619	0.0277	0.1641	Quarter 7=1, 0 otherwise
Q8	0.0280	0.1651	0.0294	0.1688	0.0280	0.1649	Quarter 8=1, 0 otherwise
Q9[#]	0.0269	0.1617	0.0306	0.1722	0.0267	0.1613	Quarter 9=1, 0 otherwise
Q10	0.0283	0.1658	0.0286	0.1667	0.0283	0.1658	Quarter 10=1, 0 otherwise
Q11	0.0286	0.1666	0.0269	0.1619	0.0286	0.1668	Quarter 11=1, 0 otherwise
Q12	0.0289	0.1674	0.0265	0.1606	0.0290	0.1677	Quarter 12=1, 0 otherwise
Q13	0.0286	0.1665	0.0292	0.1684	0.0285	0.1665	Quarter 13=1, 0 otherwise
Q14	0.0284	0.1660	0.0316	0.1750	0.0282	0.1656	Quarter 14=1, 0 otherwise
Q15	0.0277	0.1641	0.0272	0.1628	0.0277	0.1642	Quarter 15=1, 0 otherwise
Q16[#]	0.0289	0.1675	0.0254	0.1574	0.0291	0.1680	Quarter 16=1, 0 otherwise
Q17	0.0293	0.1688	0.0295	0.1692	0.0293	0.1688	Quarter 17=1, 0 otherwise
Q18	0.0296	0.1695	0.0328	0.1782	0.0294	0.1691	Quarter 18=1, 0 otherwise
Q19	0.0295	0.1691	0.0286	0.1667	0.0295	0.1692	Quarter 19=1, 0 otherwise
Q20	0.0287	0.1670	0.0281	0.1654	0.0287	0.1670	Quarter 20=1, 0 otherwise
Q21	0.0290	0.1678	0.0262	0.1597	0.0291	0.1681	Quarter 21=1, 0 otherwise
Q22	0.0297	0.1698	0.0315	0.1746	0.0296	0.1696	Quarter 22=1, 0 otherwise
Q23	0.0294	0.1689	0.0280	0.1650	0.0295	0.1691	Quarter 23=1, 0 otherwise
Q24	0.0295	0.1692	0.0272	0.1628	0.0296	0.1695	Quarter 24=1, 0 otherwise
Q25	0.0283	0.1658	0.0268	0.1615	0.0284	0.1660	Quarter 25=1, 0 otherwise
Q26	0.0288	0.1671	0.0269	0.1619	0.0288	0.1674	Quarter 26=1, 0 otherwise
Q27	0.0285	0.1664	0.0315	0.1746	0.0284	0.1660	Quarter 27=1, 0 otherwise
Q28*	0.0536	0.2252	0.0483	0.2144	0.0538	0.2257	Quarter 28=1, 0 otherwise

Q29	0.0523	0.2227	0.0525	0.2231	0.0523	0.2227	Quarter 29=1, 0 otherwise
Q30	0.0533	0.2246	0.0518	0.2216	0.0534	0.2248	Quarter 30=1, 0 otherwise
Q31	0.0516	0.2212	0.0540	0.2261	0.0515	0.2210	Quarter 31=1, 0 otherwise
N	151,944		6,608		145,336		

[1] Wage data for all workers are used in Models 1-2

[2] Separate wage data for nurses only and for all other workers only are used in Models 3-5

* Difference in mean values between nurses and all other workers significant at the 5% level

Difference in mean values between nurses and all other workers significant at the 10% level

Table A5.2.5. Descriptive statistics of variables in wage equations

Appendix 5.3. Full results of the statistical models

The tables below show the full set of results for the statistical models as discussed in the text (Section 5.7), including the time trend variables.

A5.3.1. Model 1

	β [1]	Std. Err. [2]
Constant	-0.1255*	0.0458
NURSE	0.2034*	0.0052
Years of education variables		
YED	0.1367*	0.0061
YED2	-0.0036*	0.0002
Educational attainment variables		
PGDEG	0.3900*	0.0077
DEG	0.2904*	0.0051
ALEVEL	0.0164*	0.0052
NOQUAL	-0.1427*	0.0033
Work experience variables		
EXP	0.0326*	0.0005
EXP2	-0.0005*	0.0000
Personal characteristic variables		
DISABLE	-0.0802*	0.0054
ETHNIC	-0.0151*	0.0073
NONBRIT	0.0507*	0.0079
ETHNBRIT	-0.1007*	0.0163
Regional variables		
SEAST	0.1580*	0.0026
Job characteristic variables		
HOURSPW	0.0008*	0.0001
MANAGE	0.1383*	0.0028
NWORKERS	0.1355*	0.0025
TEMP	0.0142*	0.0052
Time trend variables		
Q1	0.0231*	0.0096
Q2	0.0003	0.0100
Q3	0.0050	0.0098
Q4	0.0220*	0.0099
Q5	0.0237*	0.0097
Q6	0.0252*	0.0097
Q7	0.0073	0.0098
Q8	0.0372*	0.0097
Q9	0.0329*	0.0098
Q10	0.0257*	0.0097
Q11	0.0156	0.0097
Q12	0.0523*	0.0095

Q13	0.0443*	0.0100
Q14	0.0498*	0.0098
Q15	0.0442*	0.0098
Q16	0.0532*	0.0097
Q17	0.0577*	0.0095
Q18	0.0498*	0.0097
Q19	0.0658*	0.0094
Q20	0.0811*	0.0098
Q21	0.0841*	0.0098
Q22	0.0888*	0.0097
Q23	0.0975*	0.0096
Q24	0.1230*	0.0096
Q25	0.1372*	0.0099
Q26	0.1459*	0.0098
Q27	0.1620*	0.0097
Q28	0.1805*	0.0086
Q29	0.1830*	0.0087
Q30	0.1896*	0.0087
Q31	0.2099*	0.0087
Adjusted R^2	0.2985	
Model test	$F(49, 151{,}894) = 1320.57$; $p = 0.0000$	
N	151,944	

[1] Dependent variable is LNWAGE

[2] Results corrected for heteroscedasticity using White's estimator

* Significant at the 5% level

Table A5.3.1 (based on Table 5.3 in the main text). Results of Model 1: OLS estimates of wage equation [5.1] based on all workers with NURSE dummy

A5.3.2. Model 2

	δ [1]	Std. Err.
Constant	-1.1742*	0.0721
Age variables		
AGE	0.0748*	0.0020
AGE2	-0.0008*	0.0000
Personal characteristic variables		
DISABLE	-1.0597*	0.0094
ETHNIC	-0.4444*	0.0158
NONBRIT	-0.3276*	0.0166
ETHNBRIT	0.0101	0.0316
Family variables		
PCHILD	-0.1205*	0.0103
COHABIT	0.2329*	0.0116
MARRIED	0.0597*	0.0082
Property income variables		
PENSION	-0.5040*	0.0237
NONLABY	-0.00002*	0.0000
Years of education variables		
YED	0.2175*	0.0083
YED2	-0.0074*	0.0003

Educational attainment variables		
PGDEG	0.5698*	0.0241
DEG	0.3297*	0.0141
ALEVEL	-0.1016*	0.0123
NOQUAL	-0.5680*	0.0080
Regional variables		
SEAST	0.0061	0.0067
Time trend variables		
Q1	-1.4338*	0.0184
Q2	-1.3711*	0.0186
Q3	-2.4694*	0.0144
Q8	-2.3098*	0.0145
Q13	-2.2385*	0.0146
Q17	-1.2454*	0.0189
Q18	-1.2434*	0.0189
Q19	-1.1973*	0.0191
Q20	-1.2409*	0.0191
Q21	-1.1883*	0.0193
Q22	-1.1889*	0.0192
Q23	-1.1902*	0.0192
Q24	-1.1936*	0.0192
Q25	-1.2089*	0.0194
Q26	-1.1865*	0.0194
Q27	-1.1875*	0.0195
Q28	-1.2638*	0.0154
Q29	-1.2744*	0.0155
Q30	-1.2318*	0.0156
Q31	-1.2656*	0.0156
Log likelihood function	-115,454.20	
Restricted log likelihood	-165,334.20	
Model test	$\chi^2 = 99{,}760.03$; df = 38; sig. = 0.0000	
N	247,774	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
* Significant at the 5% level

Table A5.3.2 (based on Table 5.4 in the main text). Results of Model 2: probit estimates of participation equation [5.2] based on all individuals

	β [1]	Std. Err.	Marginal Effects	Std. Err.
Constant	0.1898*	0.0320	0.1898*	0.0320
NURSE	0.2007*	0.0057	0.2007*	0.0057
Years of education variables				
YED	0.0946*	0.0042	0.1503*	0.0067
YED2	-0.0021*	0.0001	-0.0040*	0.0002
Educational attainment variables				
PGDEG	0.2851*	0.0084	0.4310*	0.0175
DEG	0.2283*	0.0054	0.3127*	0.0105
ALEVEL	0.0423*	0.0050	0.0162#	0.0093
NOQUAL	-0.0507*	0.0045	-0.1962*	0.0068
Work experience variables				
EXP	0.0297*	0.0005	0.0297*	0.0005
EXP2	-0.0005*	0.0000	-0.0005*	0.0000

Personal characteristic variables				
DISABLE	0.1510*	0.0082	-0.1203*	0.0101
ETHNIC	0.0690*	0.0077	-0.0448*	0.0127
NONBRIT	0.1053*	0.0075	0.0214#	0.0130
ETHNBRIT	-0.0881*	0.0155	-0.0855*	0.0254
Region variables				
SEAST	0.1549*	0.0027	0.1565*	0.0050
Job characteristic variables				
HOURSPW	0.0009*	0.0001	0.0009*	0.0001
MANAGE	0.1350*	0.0027	0.1350*	0.0027
NWORKERS	0.1355*	0.0024	0.1355*	0.0024
TEMP	0.0153*	0.0043	0.0153*	0.0043
Time trend variables				
Q1	0.2333*	0.0118	-0.1338*	0.0166
Q2	0.1990*	0.0117	-0.1520*	0.0167
Q3	0.4873*	0.0165	-0.1450*	0.0189
Q4	0.0292*	0.0105	0.0292*	0.0105
Q5	0.0313*	0.0104	0.0313*	0.0104
Q6	0.0288*	0.0103	0.0288*	0.0103
Q7	0.0144	0.0103	0.0144	0.0103
Q8	0.4707*	0.0154	-0.1208*	0.0180
Q9	0.0399*	0.0104	0.0399*	0.0104
Q10	0.0288*	0.0102	0.0288*	0.0102
Q11	0.0215*	0.0103	0.0215*	0.0103
Q12	0.0535*	0.0102	0.0535*	0.0102
Q13	0.4574*	0.0150	-0.1158*	0.0176
Q14	0.0490*	0.0102	0.0490*	0.0102
Q15	0.0422*	0.0103	0.0422*	0.0103
Q16	0.0495*	0.0102	0.0495*	0.0102
Q17	0.2195*	0.0111	-0.0994*	0.0164
Q18	0.2108*	0.0111	-0.1076*	0.0164
Q19	0.2174*	0.0110	-0.0892*	0.0164
Q20	0.2389*	0.0111	-0.0788*	0.0165
Q21	0.2316*	0.0110	-0.0726*	0.0165
Q22	0.2365*	0.0110	-0.0679*	0.0164
Q23	0.2448*	0.0110	-0.0600*	0.0165
Q24	0.2702*	0.0110	-0.0354*	0.0164
Q25	0.2864*	0.0111	-0.0232	0.0166
Q26	0.2904*	0.0110	-0.0134	0.0166
Q27	0.3061*	0.0111	0.0020	0.0166
Q28	0.3380*	0.0101	0.0144	0.0141
Q29	0.3397*	0.0101	0.0134	0.0141
Q30	0.3368*	0.0100	0.0213	0.0141
Q31	0.3637*	0.0101	0.0396*	0.0142
Selection bias variables				
λ	-0.4022*	0.0110		
Adjusted R^2	0.3049			
Model test	F(50, 151,893) = 1,334.19; p = 0.0000			
N	151,944			

[1] Dependent variable is LNWAGE

* Significant at the 5% level

[#] Significant at the 10% level

Table A5.3.3 (based on Table 5.5 in the main text). Results of Model 2: participation *selection bias corrected estimates of wage equation [5.3] based on all workers with NURSE dummy*

A5.3.3. Model 3

	Nurses		All Other Workers	
	β [1]	Std. Err.[2]	β [1]	Std. Err.[2]
ONE	0.8263*	0.1498	-0.1135*	0.0462
Years of education variables				
YED	0.0975*	0.0196	0.1329*	0.0062
YED2	-0.0030*	0.0007	-0.0035*	0.0002
Educational attainment variables				
NURSEQUA	0.2389*	0.0214	0.1673*	0.0082
PGDEG	0.2252*	0.0422	0.3924*	0.0079
DEG	0.0998*	0.0164	0.2995*	0.0053
ALEVEL	-0.1683*	0.0597	0.0260*	0.0052
NOQUAL	-0.2666*	0.0904	-0.1355*	0.0033
Work experience variables				
EXP	0.0197*	0.0018	0.0327*	0.0005
EXP2	-0.0003*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	-0.0435#	0.0224	-0.0830*	0.0055
ETHNIC	-0.0262	0.0240	-0.0118	0.0076
NONBRIT	0.0659*	0.0201	0.0523*	0.0083
ETHNBRIT	0.0005	0.0423	-0.1063*	0.0172
Regional variables				
SEAST	0.0707*	0.0106	0.1613*	0.0026
Job characteristic variables				
HOURSPW	-0.0043*	0.0005	0.0009*	0.0001
MANAGE	0.0670*	0.0122	0.1365*	0.0029
NWORKERS	-0.0402*	0.0119	0.1386*	0.0025
TEMP	-0.0790*	0.0262	0.0150*	0.0053
Time trend variables				
Q1	-0.0060	0.0469	0.0243*	0.0098
Q2	-0.0160	0.0463	0.0015	0.0102
Q3	0.0315	0.0436	0.0031	0.0101
Q4	0.0004	0.0418	0.0229*	0.0101
Q5	0.0363	0.0440	0.0225*	0.0099
Q6	0.0216	0.0419	0.0246*	0.0099
Q7	0.0241	0.0438	0.0070	0.0100
Q8	0.0206	0.0441	0.0377*	0.0098
Q9	0.0053	0.0429	0.0330*	0.0100
Q10	-0.0031	0.0454	0.0272*	0.0098
Q11	0.0007	0.0427	0.0147	0.0099
Q12	0.0457	0.0467	0.0518*	0.0097
Q13	0.0805#	0.0468	0.0426*	0.0101
Q14	0.0677#	0.0408	0.0495*	0.0100
Q15	0.0180	0.0469	0.0463*	0.0099
Q16	0.0212	0.0423	0.0552*	0.0099
Q17	0.0630	0.0445	0.0580*	0.0097

Q18	0.0119	0.0478	0.0511*	0.0099
Q19	0.0336	0.0441	0.0677*	0.0096
Q20	0.0505	0.0428	0.0821*	0.0101
Q21	0.0327	0.0480	0.0842*	0.0100
Q22	0.0724#	0.0441	0.0894*	0.0099
Q23	0.0142	0.0465	0.1000*	0.0097
Q24	0.0837#	0.0430	0.1242*	0.0098
Q25	0.0961*	0.0423	0.1385*	0.0101
Q26	0.0693	0.0508	0.1481*	0.0100
Q27	0.1337*	0.0439	0.1629*	0.0098
Q28	0.1477*	0.0406	0.1817*	0.0088
Q29	0.1455*	0.0407	0.1857*	0.0089
Q30	0.1614*	0.0393	0.1901*	0.0089
Q31	0.1835*	0.0408	0.2113*	0.0089
Adjusted R^2	0.1485		0.2936	
Model test	F(49, 6,558) = 24.51; p = 0.0000		F(49, 145,285) = 1,233.62; p = 0.0000	
N	6,608		145,335	

[1] Dependent variable is LNWAGE
[2] Results corrected for heteroscedasticity using White's estimator
* Significant at the 5% level
# Significant at the 10% level

Table A5.3.4 (based on Table 5.6 in the main text). Results of Model 3: OLS estimates of separate wage equations [5.4] and [5.5] for nurses and all other workers

A5.3.4. Model 4

	Nurses		All Other Workers	
	δ [1]	Std.Err.	δ [1]	Std.Err.
Constant	1.6757*	0.5629	-1.1004*	0.0729
Age variables				
AGE	0.0536*	0.0156	0.0715*	0.0020
AGE2	-0.0009*	0.0002	-0.0008*	0.0000
Personal characteristic variables				
DISABLE	-1.2072*	0.0514	-1.0545*	0.0096
ETHNIC	0.0763	0.0950	-0.4564*	0.0161
NONBRIT	0.1245	0.0878	-0.3457*	0.0170
ETHNBRIT	-0.1727	0.1788	0.0117	0.0323
Family variables				
PCHILD	-0.0941	0.0575	-0.1297*	0.0105
COHABIT	0.3699*	0.0900	0.2316*	0.0117
MARRIED	-0.2699*	0.0466	0.0718*	0.0084
Property income variables				
PENSION	-0.8275*	0.1024	-0.4846*	0.0244
NONLABY	-0.0001*	0.0000	-0.00007*	0.0000
Years of education variables				
YED	0.0343	0.0587	0.2135*	0.0084
YED2	-0.0025	0.0019	-0.0073*	0.0003
Educational attainment				
PGDEG	0.2900#	0.1574	0.5868*	0.0245
DEG	0.0897	0.0751	0.3443*	0.0145

ALEVEL	[2]	[2]	-0.0932*	0.0123
NOQUAL	[2]	[2]	-0.5638*	0.0081
Regional variables				
SEAST	-0.1677*	0.0390	0.0125#	0.0068
Time trend variables				
Q1	-1.2981*	0.1082	-1.4393*	0.0188
Q2	-1.2531*	0.1056	-1.3750*	0.0190
Q3	-2.3712*	0.0799	-2.4738*	0.0147
Q8	-2.1562*	0.0824	-2.3131*	0.0147
Q13	-2.1280*	0.0829	-2.2404*	0.0148
Q17	-1.1443*	0.1062	-1.2460*	0.0193
Q18	-0.9645*	0.1090	-1.2507*	0.0192
Q19	-0.9798*	0.1158	-1.2013*	0.0194
Q20	-0.9028*	0.1199	-1.2477*	0.0194
Q21	-1.0490*	0.1133	-1.1886*	0.0196
Q22	-1.0811*	0.1077	-1.1917*	0.0195
Q23	-1.0088*	0.1151	-1.1944*	0.0195
Q24	-0.9368*	0.1175	-1.1977*	0.0195
Q25	-1.1191*	0.1105	-1.2084*	0.0197
Q26	-1.0142*	0.1147	-1.1882*	0.0197
Q27	-0.9732*	0.1108	-1.1936*	0.0198
Q28	-1.1393*	0.0911	-1.2652*	0.0156
Q29	-1.1312*	0.0901	-1.2768*	0.0158
Q30	-1.0146*	0.0923	-1.2355*	0.0159
Q31	-1.0701*	0.0897	-1.2689*	0.0159
Log likelihood function	-3,470.225		-111627.8	
Restricted log likelihood	-5,047.129		-159934.6	
Model test	$\chi^2 = 3,153.749$; df = 36; sig. = 0.0000		$\chi^2 = 96613.52$; df = 38; sig. = 0.0000	
N	8,878		238,896	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)

[2] ALEVEL and NOQUAL predict PART perfectly for nurses and so are omitted

* Significant at the 5% level

Significant at the 10% level

Table A5.3.5 (based on Table 5.7 in the main text). Results of Model 4: probit estimates of participation equations [5.6] and [5.8] estimated separately for nurses and all other workers

	Nurses				All Other Workers			
	β [1]	Std.Err.	Marginal Effects	Std.Err.	β [1]	Std.Err.	Marginal Effects	Std.Err.
Constant	0.8233*	0.1461	0.8233*	0.1461	0.1853*	0.0324	0.1853*	0.0324
Years of education variables								
YED	0.0979*	0.0190	0.0978*	0.0419	0.0931*	0.0042	0.1464*	0.0068
YED2	-0.0031*	0.0006	-0.0030*	0.0014	-0.0020*	0.0001	-0.0038*	0.0002
Educational attainment variables								
NURSEQUA	0.2388*	0.0186	0.2388*	0.0186	0.1608*	0.0075	0.1608*	0.0075
PGDEG	0.2257*	0.0382	0.2247*	0.1073	0.2870*	0.0086	0.4335*	0.0178
DEG	0.1000*	0.0186	0.0997*	0.0513	0.2362*	0.0056	0.3221*	0.0108
ALEVEL	-0.1681*	0.0469	-0.1681*	0.0469	0.0496*	0.0051	0.0263*	0.0093
NOQUAL	-0.2665*	0.0675	-0.2665*	0.0675	-0.0470*	0.0045	-0.1877*	0.0068
Work experience variables								
EXP	0.0197*	0.0019	0.0197*	0.0019	0.0298*	0.0005	0.0298*	0.0005

EXP2	-0.0003*	0.0001	-0.0003*	0.0001	-0.0005*	0.0000	-0.0005*	0.0000
Personal characteristic variables								
DISABLE	-0.0468	0.0298	-0.0426	0.0443	0.1419*	0.0083	-0.1213*	0.0103
ETHNIC	-0.0261	0.0244	-0.0264	0.0652	0.0738*	0.0079	-0.0401*	0.0129
NONBRIT	0.0660*	0.0220	0.0656	0.0601	0.1090*	0.0078	0.0227#	0.0133
ETHNBRIT	0.0004	0.0455	0.0010	0.1226	-0.0925*	0.0161	-0.0896*	0.0261
Regional variables								
SEAST	0.0704*	0.0106	0.0709*	0.0270	0.1572*	0.0027	0.1603*	0.0051
Job characteristic variables								
HOURSPW	-0.0043*	0.0004	-0.0043*	0.0004	0.0010*	0.0001	0.0010*	0.0001
MANAGE	0.0670*	0.0110	0.0670*	0.0110	0.1333*	0.0028	0.1333*	0.0028
NWORKERS	-0.0400*	0.0121	-0.0400*	0.0121	0.1386*	0.0024	0.1386*	0.0024
TEMP	-0.0791*	0.0190	-0.0791*	0.0190	0.0157*	0.0044	0.0157*	0.0044
Time trend variables								
Q1	-0.0079	0.0408	-0.0034	0.0801	0.2317*	0.0121	-0.1275*	0.0170
Q2	-0.0178	0.0407	-0.0134	0.0786	0.1971*	0.0120	-0.1461*	0.0170
Q3	0.0262	0.0513	0.0345	0.0723	0.4765*	0.0169	-0.1409*	0.0193
Q4	0.0003	0.0395	0.0003	0.0395	0.0299*	0.0107	0.0299*	0.0107
Q5	0.0362	0.0386	0.0362	0.0386	0.0296*	0.0106	0.0296*	0.0106
Q6	0.0216	0.0381	0.0216	0.0381	0.0277*	0.0105	0.0277*	0.0105
Q7	0.0240	0.0392	0.0240	0.0392	0.0136	0.0105	0.0136*	0.0105
Q8	0.0160	0.0487	0.0235	0.0715	0.4630*	0.0158	-0.1143*	0.0184
Q9	0.0051	0.0381	0.0051	0.0381	0.0395*	0.0106	0.0395*	0.0106
Q10	-0.0031	0.0386	-0.0031	0.0386	0.0301*	0.0104	0.0301*	0.0104
Q11	0.0006	0.0392	0.0006	0.0392	0.0201*	0.0104	0.0201*	0.0104
Q12	0.0456	0.0393	0.0456	0.0393	0.0527*	0.0104	0.0527*	0.0104
Q13	0.0762	0.0481	0.0836	0.0714	0.4473*	0.0154	-0.1118*	0.0180
Q14	0.0676#	0.0378	0.0676	0.0378	0.0480*	0.0104	0.0480*	0.0104
Q15	0.0180	0.0391	0.0180	0.0391	0.0441*	0.0105	0.0441*	0.0105
Q16	0.0211	0.0398	0.0211	0.0398	0.0512*	0.0104	0.0512*	0.0104
Q17	0.0613	0.0401	0.0653	0.0786	0.2165*	0.0113	-0.0945*	0.0167
Q18	0.0105	0.0386	0.0139	0.0794	0.2099*	0.0113	-0.1022*	0.0167
Q19	0.0323	0.0397	0.0357	0.0837	0.2169*	0.0112	-0.0829*	0.0167
Q20	0.0493	0.0396	0.0525	0.0860	0.2379*	0.0114	-0.0734*	0.0168
Q21	0.0313	0.0408	0.0349	0.0828	0.2287*	0.0112	-0.0679*	0.0168
Q22	0.0709#	0.0393	0.0746	0.0790	0.2342*	0.0112	-0.0632*	0.0167
Q23	0.0129*	0.0400	0.0164	0.0835	0.2449*	0.0112	-0.0532*	0.0167
Q24	0.0825*	0.0400	0.0857	0.0849	0.2689*	0.0112	-0.0299#	0.0167
Q25	0.0945*	0.0408	0.0984	0.0813	0.2844*	0.0113	-0.0172	0.0169
Q26	0.0679#	0.0405	0.0714	0.0835	0.2898*	0.0113	-0.0068	0.0169
Q27	0.1325*	0.0390	0.1358#	0.0806	0.3049*	0.0113	0.0070	0.0169
Q28	0.1461*	0.0369	0.1500*	0.0687	0.3359*	0.0103	0.0201	0.0143
Q29	0.1439*	0.0364	0.1478*	0.0679	0.3394*	0.0103	0.0208	0.0144
Q30	0.1600*	0.0362	0.1636*	0.0690	0.3346*	0.0102	0.0263#	0.0144
Q31	0.1820*	0.0362	0.1857*	0.0676	0.3622*	0.0104	0.0456*	0.0145
Selection bias variables								
λ	0.0055	0.0361			-0.3920*	0.0112		
Adjusted R^2	0.1483				0.2997			
Model test	F(50, 6,557) = 24.02; p value = 0.0000				F(50, 145285)= 1245.23; p value = 0.0000			
N	6,608				145,336			

[1] Dependent variable is LNWAGE

* Significant at the 5% level

\# Significant at the 10% level

Table A5.3.6 (based on Table 5.8 in the main text). Results of Model 4: participation selection bias corrected estimates of wage equations [5.7] and [5.9] estimated separately for nurses and all other workers

A5.3.5. Model 5

	γ^{1}	Std. Err.
Constant	-2.7510*	0.3250
NURSEQUA	3.0550*	0.0227
Age variables		
AGE	0.0012	0.0074
AGE2	-0.0001	0.0001
Personal characteristic variables		
DISABLE	-0.1554*	0.0470
ETHNIC	0.2124*	0.0537
NONBRIT	0.3157*	0.0501
ETHNBRIT	0.0562	0.1037
Family variables		
PCHILD	0.0937*	0.0321
COHABIT	-0.0666#	0.0376
MARRIED	-0.1150*	0.0274
Property income variables		
PENSION	-0.0404	0.0964
NONLABY	-0.00003*	0.0000
Years of education variables		
YED	0.0452	0.0388
YED2	-0.0022#	0.0013
Educational attainment variables		
PGDEG	-0.5472*	0.0738
DEG	-0.0804*	0.0398
ALEVEL	0.1444*	0.0480
NOQUAL	-0.3265*	0.0598
Region variables		
SEAST	-0.0980*	0.0227
Time trend variables		
Q1	0.0842	0.0865
Q2	0.1089	0.0878
Q3	0.1231	0.0866
Q4	0.0271	0.0883
Q5	0.0237	0.0872
Q6	0.0827	0.0864
Q7	0.0590	0.0879
Q8	0.0739	0.0868
Q9	0.0539	0.0872
Q10	0.0377	0.0869
Q11	-0.0922	0.0887
Q12	-0.0340	0.0884
Q13	0.0842	0.0864
Q14	0.1883*	0.0847
Q15	0.1100	0.0869

Q16	0.0356	0.0885
Q17	0.1000	0.0862
Q18	0.1048	0.0845
Q19	0.1511#	0.0850
Q20	0.0457	0.0871
Q21	-0.0754	0.0888
Q22	0.0965	0.0853
Q23	-0.0170	0.0868
Q24	-0.0101	0.0875
Q25	0.0214	0.0875
Q26	-0.0516	0.0880
Q27	0.1345	0.0848
Q28	0.0527	0.0771
Q29	0.1864*	0.0759
Q30	0.1194	0.0765
Q31	0.1786*	0.0758
Log likelihood function	-9,528.88	
Restricted log likelihood	27,179.79	
Model test	$\chi^2 = 35{,}301.81$; df = 50; sig. = 0.0000	
N	151,944	

[1] Dependent variable is whether the individual is employed as a nurse (NURSE = 1) or not (NURSE = 0).
* Significant at the 5% level
# Significant at the 10% level

Table A5.3.7 (based on Table 5.9 in the main text). Results of Model 5: probit estimates of occupational selection equation [5.10] based on all workers

	Nurses				All Other Workers			
	β [1]	Std. Err.	Marginal Effects	Std. Err.	β [1]	Std. Err.	Marginal Effects	Std. Err.
ONE	0.2203	0.3002	0.2203	0.3002	-0.1142*	0.0299	-0.1142*	0.0299
Years of education variables								
YED	0.1100*	0.0199	0.1050*	0.0317	0.1332*	0.0039	0.1315*	0.0250
YED2	-0.0035*	0.0007	-0.0032*	0.0010	-0.0035*	0.0001	-0.0034*	0.0008
Educational attainment variables								
NURSEQUA	0.6542*	0.1803	0.3144#	0.1803	0.2240*	0.0520	0.1126*	0.0520
PGDEG	0.1587*	0.0481	0.2196*	0.0673	0.3911*	0.0073	0.4110*	0.0476
DEG	0.0920*	0.0192	0.1010*	0.0318	0.2992*	0.0048	0.3021*	0.0258
ALEVEL	-0.1415*	0.0470	-0.1576*	0.0560	0.0263*	0.0046	0.0210	0.0309
NOQUAL	-0.3274*	0.0703	-0.2910*	0.0799	-0.1360*	0.0035	-0.1241*	0.0382
Work experience variables								
EXP	0.0194*	0.0019	0.0194*	0.0019	0.0327*	0.0005	0.0327*	0.0005
EXP2	-0.0003*	0.0001	-0.0003*	0.0001	-0.0005*	0.0000	-0.0005*	0.0000
Personal characteristic variables								
DISABLE	-0.0612*	0.0224	-0.0439	0.0374	-0.0834*	0.0051	-0.0777*	0.0303
ETHNIC	-0.0073	0.0263	-0.0309	0.0431	-0.0112	0.0072	-0.0189	0.0349
NONBRIT	0.0967*	0.0263	0.0616	0.0413	0.0533*	0.0072	0.0418	0.0327
ETHNBRIT	0.0079	0.0468	0.0016	0.0809	-0.1059*	0.0156	-0.1080	0.0678
Regional variables								
SEAST	0.0592*	0.0117	0.0701*	0.0186	0.1611*	0.0025	0.1646*	0.0147
Job characteristic variables								
HOURSPW	-0.0042*	0.0004	-0.0042*	0.0004	0.0009*	0.0001	0.0009*	0.0001
MANAGE	0.0672*	0.0110	0.0672*	0.0110	0.1365*	0.0028	0.1365*	0.0028

NWORKERS	-0.0363*	0.0121	-0.0363*	0.0121	0.1386*	0.0024	0.1386*	0.0024
TEMP	-0.0785*	0.0190	-0.0785*	0.0190	0.0149*	0.0044	0.0149*	0.0044
Time trend variables								
Q1	0.0015	0.0399	-0.0078	0.0680	0.0245*	0.0097	0.0215	0.0559
Q2	-0.0049	0.0401	-0.0170	0.0688	0.0017	0.0098	-0.0023	0.0567
Q3	0.0434	0.0394	0.0297	0.0678	0.0034	0.0098	-0.0011	0.0560
Q4	0.0025	0.0405	-0.0005	0.0693	0.0229*	0.0098	0.0220	0.0571
Q5	0.0390	0.0396	0.0364	0.0682	0.0225*	0.0098	0.0217	0.0563
Q6	0.0294	0.0392	0.0202	0.0675	0.0248*	0.0097	0.0218	0.0558
Q7	0.0292	0.0402	0.0226	0.0689	0.0071	0.0097	0.0050	0.0568
Q8	0.0261	0.0394	0.0178	0.0679	0.0379*	0.0096	0.0352	0.0561
Q9	0.0093	0.0391	0.0033	0.0679	0.0330*	0.0098	0.0311	0.0564
Q10	0.0005	0.0396	-0.0037	0.0680	0.0272*	0.0096	0.0258	0.0561
Q11	-0.0104	0.0404	-0.0001	0.0695	0.0144	0.0096	0.0178	0.0573
Q12	0.0401	0.0404	0.0439	0.0692	0.0517*	0.0096	0.0529	0.0571
Q13	0.0869*	0.0395	0.0775	0.0677	0.0427*	0.0096	0.0396	0.0558
Q14	0.0837*	0.0394	0.0627	0.0667	0.0499*	0.0096	0.0430	0.0548
Q15	0.0274	0.0402	0.0151	0.0684	0.0465*	0.0097	0.0425	0.0561
Q16	0.0224	0.0408	0.0184	0.0696	0.0552*	0.0096	0.0539	0.0572
Q17	0.0696#	0.0395	0.0585	0.0676	0.0581*	0.0096	0.0545	0.0557
Q18	0.0199	0.0387	0.0083	0.0663	0.0512*	0.0096	0.0474	0.0546
Q19	0.0457	0.0400	0.0289	0.0673	0.0679*	0.0096	0.0624	0.0550
Q20	0.0516	0.0398	0.0466	0.0683	0.0821*	0.0096	0.0805	0.0563
Q21	0.0211	0.0408	0.0295	0.0697	0.0839*	0.0096	0.0866	0.0573
Q22	0.0792*	0.0391	0.0684	0.0669	0.0895*	0.0096	0.0860	0.0552
Q23	0.0086	0.0400	0.0105	0.0682	0.0998*	0.0096	0.1004*	0.0561
Q24	0.0788*	0.0402	0.0799	0.0687	0.1240*	0.0096	0.1244*	0.0565
Q25	0.0936*	0.0403	0.0912	0.0688	0.1384*	0.0097	0.1376*	0.0566
Q26	0.0597	0.0405	0.0654	0.0691	0.1478*	0.0097	0.1497*	0.0568
Q27	0.1435*	0.0392	0.1286#	0.0668	0.1631*	0.0097	0.1582*	0.0549
Q28	0.1493*	0.0361	0.1434*	0.0609	0.1816*	0.0086	0.1797*	0.0498
Q29	0.1603*	0.0363	0.1396*	0.0604	0.1860*	0.0086	0.1792*	0.0491
Q30	0.1687*	0.0359	0.1555*	0.0359	0.1902*	0.0086	0.1858*	0.0086
Q31	0.1969*	0.0360	0.1770*	0.0360	0.2116*	0.0087	0.2051*	0.0087
Selection bias variables								
λ	0.1747*	0.0755			-0.0573	0.0520		
Adjusted R^2	0.1490				0.2936			
Model test	$F(50, 6{,}557) = 24.14$; $p = 0.0000$				$F(50, 145{,}284) = 1{,}208.98$; $p = 0.0000$			
N	6,608				145,335			

[1] Dependent variable is LNWAGE
* Significant at the 5% level
\# Significant at the 10% level

Table A5.3.8 (based on Table 5.10 in the main text). Results of Model 5: occupation selection bias corrected wage equations [5.11] and [5.13] estimated separately for nurses and all other workers

Appendix 5.4. A note on the use of semi-logarithmic models

Semi-log(arithmic) models of the type

$$\ln Y_i = \beta_0 + \beta_1 X_i + U_i \qquad [A5.4.1]$$

have a useful property in that when the independent variable X is continuous its co-efficient β_1 measures the constant proportional or relative change in Y for a given absolute (say, unitary) change in X. That is:

$$\beta_1 = \frac{d(\ln Y)}{dX} = \frac{1}{Y}\frac{dY}{dX} = \frac{(dY/Y)}{dX} = \frac{\text{relative change in Y}}{\text{absolute change in X}} \qquad [A5.4.2]$$

If we multiply the relative change in Y by 100, equation [A5.4.2] then measures the percentage change in Y for an absolute change in continuous variable X. In the present analysis this semi-log form is adopted and the earnings functions we estimate are of the type:

$$\ln W_i = \beta X_i + U_i \qquad [A5.4.3]$$

From equation [A5.4.2] the co-efficients (the β's) may thus be interpreted as the constant proportional change in wages W for a given absolute change in continuous variable X.

The same interpretation is not possible, however, when the independent variable is a dummy variable and enters the equation in dichotomous form. This is because the derivative of the dependent variable lnY with respect to the dummy variable does not exist. Instead the co-efficient of the dummy variable measures the discontinuous effect on Y of the presence of the factor represented by the dummy variable. The appropriate interpretation of the co-efficient of a dummy variable is given below (the following is adapted from Halvorsen and Palmqvist, 1980).

Suppose the following semi-log model:

$$\ln Y_i = \beta_0 + \beta_1 X_i + cD_i + U_i \qquad \text{[A5.4.4]}$$

where D is a dummy variable taking the value zero or one and c is its co-efficient. Depending on the value of D we therefore have the following estimated values of lnY:

$$\ln Y = \beta_0 + \beta_1 X_i + c \text{ when } D = 1\text{; call this } \ln Y_d \qquad \text{[A5.4.5a]}$$

$$\ln Y = \beta_0 + \beta_1 X_i \text{ when } D = 0\text{; call this } \ln Y_o \qquad \text{[A5.4.5b]}$$

From this we have the following:

$$\ln Y_d - \ln Y_o = \ln\left(\frac{Y_d}{Y_o}\right) = c \qquad \text{[A5.4.6]}$$

In other words c gives us the change in lnY with the presence of the factor represented by the dummy variable. However, we in fact wish to know the proportionate change in Y for a change in D. Let us call this g. In other words:

$$g = \frac{Y_d - Y_o}{Y_o} \qquad [A5.4.7]$$

This may be rearranged as follows:

$$g = \frac{Y_d}{Y_o} - 1 \qquad [A5.4.8]$$

$$\therefore \frac{Y_d}{Y_o} = g + 1 \qquad [A5.4.9]$$

$$\therefore \ln\left(\frac{Y_d}{Y_o}\right) = \ln(g + 1) \qquad [A5.4.10]$$

Combining equations [A5.4.6] and [A5.4.10] we now have:

$$c = \ln(g + 1) \qquad [A5.4.11]$$

To obtain the relative effect g on Y of the dummy variable D given its co-efficient c we must therefore calculate the following:

$$g = e^c - 1 \qquad [A5.4.12]$$

i.e. take the antilog of c and subtract one. The percentage effect is therefore equal to:

$$100*g = [e^c - 1]*100 \qquad [A5.4.13]$$

In simple terms, from the estimated model we know c but we in fact wish to obtain g. It is therefore necessary to transform the results of the original model as described above. For small values of g, c is approximately equal to c (intuitively, this is based on the approximation that $\ln[1 + x] \approx x$ when x is small). When g is positive, c is smaller than g, and when g is negative c is algebraically smaller than g but larger in absolute value.

Appendix 5.5. A note on the use of quadratic regression models

Quadratic regression models of the type:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X^2_i + U_i \qquad \text{[A5.5.1]}$$

which represents a second-degree polynomial in the variable X, are useful in statistical estimation of economic models because they allow estimation of a non-linear (parabola) relationship between X and Y. Specifically in the case of the quadratic function this model allows for a u-shaped or n-shaped relationship between X and Y. Formally, we differentiate equation [A5.5.1] twice to obtain the following:

$$\frac{dY}{dX} = \beta_1 + 2\beta_2 X \qquad \text{[A5.5.2]}$$

$$\frac{d^2Y}{dX^2} = 2\beta_2 \qquad \text{[A5.5.3]}$$

At the turning point,

$$\frac{dY}{dX} = 0 \Rightarrow \beta_1 + 2\beta_2 X = 0 \Rightarrow X = \frac{\beta_1}{2\beta_2} \qquad \text{[A5.5.4]}$$

If $\frac{d^2Y}{dX^2}$ is negative the relationship between X and Y is n-shaped with the maximum value of Y occurring where $X = \frac{\beta_1}{2\beta_2}$; if $\frac{d^2Y}{dX^2}$ is positive the relationship between X and Y is u-shaped with the minimum value of Y occurring where $X = \frac{\beta_1}{2\beta_2}$.

The quadratic regression model is relevant to the present analysis since quadratic terms are included on the years of education variables and work experience variables in the wage equations to capture non-linearities in the relationship between these variables and earnings. This is consistent with the Mincerian earnings function model (Mincer, 1974) presented in Chapter 4.

Appendix 5.6. Identification of the wage equations in Models 2, 4 and 5

Below we distinguish the identifying variables of the wage equations for Models 2, 4 and 5 (those which utilise the Heckman two-step procedure). The participation or occupation selection equation is estimated as described in the text. The wage equation however is re-estimated in each instance with no exclusion restrictions. One or both of the property income variables (PENSION or NONLABY) is in each instance found to be statistically insignificant at conventional levels.

	β [1]	Std. Err.
Constant	-1.2165*	0.0390
NURSE	0.1905*	0.0055
Years or education variables		
YED	0.1466*	0.0042
YED2	-0.0036*	0.0001
Educational attainment variables		
PGDEG	0.3255*	0.0077
DEG	0.2521*	0.0050
ALEVEL	0.0384*	0.0046
NOQUAL	-0.1640*	0.0050
Work experience variables		
EXP	0.0244*	0.0005
EXP2	-0.0003*	0.00002
Personal characteristic variables		
DISABLE	-0.1080*	0.0099
ETHNIC	-0.0212*	0.0074
NONBRIT	0.0323*	0.0071
ETHNBRIT	-0.1202*	0.0147
Regional variables		
SEAST	0.1531*	0.0024
Job characteristic variables		
HOURSPW	0.0014*	0.0001
MANAGE	0.1219*	0.0027
NWORKERS	0.1315*	0.0023
TEMP	0.0152*	0.0043
Time trend variables		
Q1	-0.0015	0.0121
Q2	-0.0233#	0.0119
Q3	-0.0533*	0.0201
Q4	0.0169#	0.0094
Q5	0.0156#	0.0094
Q6	0.0173#	0.0093
Q7	-0.0023	0.0093
Q8	-0.0226	0.0186
Q9	0.0204*	0.0094

Q10	0.0129	0.0092
Q11	-0.0026	0.0092
Q12	0.0343*	0.0092
Q13	-0.0196	0.0180
Q14	0.0323*	0.0092
Q15	0.0242*	0.0093
Q16	0.0316*	0.0092
Q17	0.0157	0.0111
Q18	0.0077	0.0111
Q19	0.0222*	0.0110
Q20	0.0327*	0.0112
Q21	0.0369*	0.0110
Q22	0.0399*	0.0109
Q23	0.0496*	0.0109
Q24	0.0737*	0.0110
Q25	0.0837*	0.0111
Q26	0.0922*	0.0110
Q27	0.1064*	0.0110
Q28	0.1184*	0.0105
Q29	0.1220*	0.0105
Q30	0.1278*	0.0104
Q31	0.1465*	0.0105
Selection bias variables		
λ	0.0447*	0.0148
Age variables		
AGE	0.0513*	0.0009
AGE2	-0.0006*	0.0000
Family variables		
PCHILD	0.0755*	0.0038
COHABIT	0.0655*	0.0042
MARRIED	-0.0162*	0.0031
Property income variables		
PENSION	0.0739*	0.0110
NONLABY	0.000002	0.000001
Adjusted R^2	0.3256	
Model test	$F(57, 151{,}886) = 1{,}228.13$; $p = 0.0000$	
N	151,944	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
[#] Significant at the 10% level

Table A.5.6.1. Results of the wage equation for Model 2 with no exclusion restrictions

	Nurses		All Other Workers	
	β [1]	Std.Err.	β [1]	Std.Err.
Constant	0.0365	0.1712	-1.1847*	0.0394
Years of education variables				
YED	0.1129*	0.0192	0.1423*	0.0042
YED2	-0.0034*	0.0006	-0.0035*	0.0001
Educational attainment variables				
NURSEQUA	0.2117*	0.0187	0.1419*	0.0073
PGDEG	0.2026*	0.0383	0.3275*	0.0079

DEG	0.0892*	0.0186	0.2598*	0.0051
ALEVEL	-0.1452*	0.0467	0.0460*	0.0046
NOQUAL	-0.2953*	0.0672	-0.1556*	0.0050
Work experience variables				
EXP	0.0167*	0.0019	0.0248*	0.0005
EXP2	-0.0002*	0.0001	-0.0003*	0.00001
Personal characteristic variables				
DISABLE	-0.0603	0.0420	-0.1088*	0.0100
ETHNIC	-0.0402#	0.0244	-0.0176*	0.0077
NONBRIT	0.0680*	0.0221	0.0327*	0.0074
ETHNBRIT	0.0023	0.0454	-0.1266*	0.0153
Regional variables				
SEAST	0.0663*	0.0109	0.1567*	0.0025
Job characteristic variables				
HOURSPW	-0.0038*	0.0004	0.0016*	0.0001
MANAGE	0.0636*	0.0110	0.1207*	0.0027
NWORKERS	-0.0279*	0.0121	0.1341*	0.0024
TEMP	-0.0730*	0.0189	0.0162*	0.0044
Time trend variables				
Q1	-0.0154	0.0444	0.0006	0.0125
Q2	-0.0241	0.0442	-0.0215#	0.0123
Q3	0.0016	0.0718	-0.0525*	0.0206
Q4	-0.0022	0.0393	0.0180#	0.0096
Q5	0.0339	0.0384	0.0147	0.0096
Q6	0.0176	0.0379	0.0170#	0.0095
Q7	0.0182	0.0390	-0.0022	0.0095
Q8	-0.0084	0.0650	-0.0197	0.0190
Q9	-0.0072	0.0380	0.0212*	0.0096
Q10	-0.0072	0.0385	0.0145	0.0095
Q11	-0.0134	0.0391	-0.0027	0.0094
Q12	0.0357	0.0391	0.0343*	0.0094
Q13	0.0477	0.0639	-0.0187	0.0184
Q14	0.0565	0.0377	0.0323*	0.0095
Q15	0.0119	0.0389	0.0263*	0.0095
Q16	0.0111	0.0396	0.0341*	0.0094
Q17	0.0405	0.0427	0.0175	0.0114
Q18	-0.0053	0.0401	0.0103	0.0114
Q19	0.0093	0.0413	0.0256	0.0112
Q20	0.0229	0.0409	0.0354*	0.0115
Q21	0.0055	0.0428	0.0390*	0.0112
Q22	0.0512	0.0414	0.0417*	0.0112
Q23	-0.0103	0.0418	0.0538*	0.0112
Q24	0.0609	0.0414	0.0765*	0.0112
Q25	0.0649	0.0432	0.0870*	0.0114
Q26	0.0394	0.0423	0.0965*	0.0113
Q27	0.1056*	0.0408	0.1089*	0.0113
Q28	0.1203*	0.0398	0.1213*	0.0107
Q29	0.1098*	0.0394	0.1266*	0.0108
Q30	0.1293*	0.0384	0.1301*	0.0106
Q31	0.1483*	0.0388	0.1499*	0.0108
Selection bias variables				
λ	0.0253	0.0636	0.0425*	0.0151
Age variables				
AGE	0.0326*	0.0044	0.0507*	0.0009

AGE2	-0.0004*	0.0001	-0.0006*	0.0000
Family variables				
PCHILD	0.0445*	0.0143	0.0763*	0.0039
COHABIT	0.0471*	0.0181	0.0654*	0.0043
MARRIED	-0.0217	0.0135	-0.0171*	0.0032
Property income variables				
PENSION	-0.0147	0.0476	0.0762*	0.0112
NONLABY	0.000006	0.000005	0.000001	0.000001
Adjusted R^2	0.1583		0.3204	
Model test	F(57, 6,550) = 22.80; p value = 0.0000		F(57, 145,278)= 1,203.21; p value = 0.0000	
N	6,608		145,336	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
[#] Significant at the 10% level

Table A.5.6.2. Results of the wage equations for Model 4 with no exclusion restrictions

	Nurses		All Other Workers	
	β [1]	Std. Err.	β [1]	Std. Err.
Constant	1.4660*	0.4787	-1.1229*	0.0327
Years of education variables				
YED	0.0982*	0.0220	0.1371*	0.0039
YED2	-0.0027*	0.0007	-0.0033*	0.0001
Educational attainment variables				
NURSEQUA	-0.9137*	0.3521	-0.0149	0.0508
PGDEG	0.3743*	0.0699	0.3217*	0.0073
DEG	0.1105*	0.0224	0.2551*	0.0048
ALEVEL	-0.2136*	0.0418	0.0471*	0.0046
NOQUAL	-0.1438*	0.0673	-0.1442*	0.0034
Work experience variables				
EXP	0.0168*	0.0019	0.0247*	0.0005
EXP2	-0.0002*	0.0001	-0.0003*	0.00001
Personal characteristic variables				
DISABLE	-0.0027	0.0276	-0.0832*	0.0051
ETHNIC	-0.0991*	0.0343	-0.0104	0.0071
NONBRIT	-0.0201	0.0390	0.0364*	0.0071
ETHNBRIT	-0.0053	0.0538	-0.1258*	0.0153
Regional variables				
SEAST	0.0953*	0.0150	0.1572*	0.0025
Job characteristic variables				
HOURSPW	-0.0038*	0.0004	0.0016*	0.0001
MANAGE	0.0642*	0.0108	0.1206*	0.0027
NWORKERS	-0.0277*	0.0120	0.1340*	0.0024
TEMP	-0.0745*	0.0185	0.0162*	0.0044
Time trend variables				
Q1	-0.0308	0.0457	0.0227*	0.0096
Q2	-0.0455	0.0463	-0.0008	0.0096
Q3	-0.0074	0.0458	-0.0018	0.0096
Q4	-0.0083	0.0459	0.0189*	0.0096
Q5	0.0281	0.0450	0.0157	0.0096
Q6	-0.0057	0.0451	0.0171[#]	0.0095

Q7	0.0023	0.0458	-0.0014	0.0095
Q8	-0.0074	0.0452	0.0265*	0.0095
Q9	-0.0222	0.0449	0.0222*	0.0096
Q10	-0.0179	0.0451	0.0152	0.0095
Q11	0.0115	0.0462	-0.0007	0.0094
Q12	0.0451	0.0458	0.0354*	0.0094
Q13	0.0458	0.0453	0.0255*	0.0095
Q14	0.0056	0.0471	0.0317*	0.0095
Q15	-0.0189	0.0464	0.0261*	0.0095
Q16	0.0000	0.0463	0.0345*	0.0094
Q17	0.0208	0.0456	0.0351*	0.0094
Q18	-0.0285	0.0447	0.0280*	0.0094
Q19	-0.0262	0.0468	0.0420*	0.0094
Q20	0.0142	0.0454	0.0534*	0.0095
Q21	0.0320	0.0464	0.0566*	0.0095
Q22	0.0316	0.0450	0.0583*	0.0094
Q23	-0.0007	0.0452	0.0711*	0.0095
Q24	0.0672	0.0455	0.0938*	0.0095
Q25	0.0648	0.0458	0.1042*	0.0096
Q26	0.0600	0.0460	0.11398	0.0095
Q27	0.0737	0.0460	0.1251*	0.0096
Q28	0.1117*	0.0411	0.1396*	0.0085
Q29	0.0650	0.0439	0.1441*	0.0085
Q30	0.1023*	0.0419	0.1472*	0.0085
Q31	0.1056*	0.0435	0.1672*	0.0086
Selection bias variables				
λ	-0.4703*	0.1466	0.1583*	0.0506
Age variables				
AGE	0.0319*	0.0045	0.0494*	0.0008
AGE2	-0.0003*	0.0001	-0.0006*	0.000009
Family variables				
PCHILD	0.0207	0.0183	0.0775*	0.0039
COHABIT	0.0638*	0.0210	0.0618*	0.0041
MARRIED	0.0118	0.0174	-0.0175*	0.0032
Property income variables				
PENSION	0.0083	0.0488	0.0860*	0.0107
NONLABY	0.00001	0.000007	0.000002	0.000001
Adjusted R^2	0.1590		0.3204	
Model test	F(57, 6,550) = 22.92; p = 0.0000		F(57, 145,278) = 1,203.05; p = 0.0000	
N	6,608		145,336	

[1] Dependent variable is LNWAGE

* Significant at the 5% level

[#] Significant at the 10% level

Table A.5.6.3. Results of the wage equations for Model 5 with no exclusion restrictions

Appendix 5.7. Model 4a

Model 4a is a modified version of the participation selection bias model described in Model 4. Separate wage equations are estimated for nurses and other workers, with a correction for participation selection bias using the Heckman two-step procedure. The difference is that in Model 4 separate participation equations to address the problem of participation selection bias are estimated. In Model 4a it is assumed that the participation equation is the same for all individuals. In the text this restriction was rejected using a Chow-type test based on the likelihood ratio statistic. For information however the full specification and results of Model 4a are presented here.

For individuals employed as nurses the following model is estimated:

Participation equation: $P_i^* = \delta Z_i + V_i$ [A5.7.1]

Wage equation: $\ln W_{ni} = \beta_n X_{ni} + \beta_{\lambda(p)n}\lambda(p)_{ni} + \varepsilon_{ni}$ [A5.7.2]

The participation equation is estimated using the whole sample of individuals in the data (participators and non-participators). Therefore $\lambda(p)$ is estimated for each observation in the whole sample. The wage equation is estimated using the sub-sample of individuals who are employed as nurses.

For individuals employed in occupations other than nursing the following model is estimated:

Participation equation: $P_i^* = \delta Z_i + V_i$ [A5.7.3]

Wage equation: $\ln W_{oi} = \beta_o X_{oi} + \beta_{\lambda(p)o}\lambda(p)_{oi} + \varepsilon_{oi}$ [A5.7.4]

The participation equation is again estimated using the whole sample of individuals in the data (participators and non-participators). Once again, $\lambda(p)$ is estimated for each observation in the whole sample. The wage equation is estimated using the sub-sample of those who are employed in occupations other than nursing.

Note that the participation equations delineated in equations [A5.7.1] and [A5.7.3] will yield identical results in terms of the co-efficients δ, and these will be identical to those in equation [5.2] pertaining to Model 2, in the text.

The results of the wage equations for Model 4a are presented in Table A5.7.1.

	Nurses				All Other Workers			
	β^{1}	Std. Err.	Marginal Effects	Std. Err.	β^{1}	Std. Err.	Marginal Effects	Std. Err.
ONE	0.9767*	0.1463	0.9767*	0.1463	0.1971*	0.0325	0.1971*	0.0325
Years of education variables								
YED	0.0775*	0.0190	0.1103*	0.0197	0.0914*	0.0042	0.1462*	0.0068
YED2	-0.0023*	0.0006	-0.0034*	0.0007	-0.0020*	0.0001	-0.0038*	0.0002
Educational attainment variables								
NURSEQUA	0.2323*	0.0185	0.2323*	0.0185	0.1610*	0.0075	0.1610*	0.0075
PGDEG	0.1716*	0.0407	0.2577*	0.0435	0.2886*	0.0086	0.4322*	0.0176
DEG	0.0674*	0.0202	0.1172*	0.0221	0.2378*	0.0055	0.3209*	0.0106
ALEVEL	-0.1581*	0.0466	-0.1734*	0.0473	0.0512*	0.0051	0.0256*	0.0093
NOQUAL	-0.2152*	0.0666	-0.3011*	0.0668	-0.0452*	0.0045	-0.1883*	0.0068
Work experience variables								
EXP	0.0189*	0.0019	0.0189*	0.0019	0.0298*	0.0005	0.0298*	0.0005
EXP2	-0.0003*	0.0001	-0.0003*	0.0001	-0.0005*	0.0000	-0.0005*	0.0000
Personal characteristic variables								
DISABLE	0.0851*	0.0354	-0.0751*	0.0359	0.1452*	0.0083	-0.1219*	0.0103
ETHNIC	0.0139	0.0259	-0.0533#	0.0277	0.0717*	0.0079	-0.0403*	0.0128
NONBRIT	0.0992*	0.0231	0.0497*	0.0254	0.1058*	0.0077	0.0233#	0.0131
ETHNBRIT	0.0063	0.0444	0.0079	0.0487	-0.0931*	0.0160	-0.0905*	0.0257
Regional variables								
SEAST	0.0688*	0.0104	0.0697*	0.0112	0.1583*	0.0027	0.1599*	0.0051

Job characteristic variables								
HOURSPW	-0.0042*	0.0004	-0.0042*	0.0004	0.0010*	0.0001	0.0010*	0.0001
MANAGE	0.0672*	0.0110	0.0672*	0.0110	0.1332*	0.0028	0.1332*	0.0028
NWORKERS	-0.0360*	0.0121	-0.0360*	0.0121	0.1384*	0.0024	0.1384*	0.0024
TEMP	-0.0770*	0.0189	-0.0770*	0.0189	0.0160*	0.0044	0.0160*	0.0044
Time trend variables								
Q1	0.1064*	0.0470	-0.1105*	0.0484	0.2320*	0.0120	-0.1293*	0.0168
Q2	0.0920*	0.0464	-0.1153*	0.0479	0.1976*	0.0119	-0.1479*	0.0168
Q3	0.3004*	0.0717	-0.0730	0.0723	0.4790*	0.0168	-0.1432*	0.0192
Q4	0.0036	0.0414	0.0036	0.0414	0.0300*	0.0107	0.0300*	0.0107
Q5	0.0419	0.0405	0.0419	0.0405	0.0298*	0.0106	0.0298*	0.0106
Q6	0.0247	0.0399	0.0247	0.0399	0.0279*	0.0105	0.0279*	0.0105
Q7	0.0286	0.0411	0.0286	0.0411	0.0138	0.0105	0.0138	0.0105
Q8	0.2599*	0.0664	-0.0894	0.0670	0.4655*	0.0157	-0.1165*	0.0182
Q9	0.0082	0.0400	0.0082	0.0400	0.0398*	0.0106	0.0398*	0.0106
Q10	-0.0032	0.0405	-0.0032	0.0405	0.0303*	0.0105	0.0303*	0.0105
Q11	0.0060	0.0412	0.0060	0.0412	0.0204#	0.0105	0.0204#	0.0105
Q12	0.0466	0.0412	0.0466	0.0412	0.0530*	0.0104	0.0530*	0.0104
Q13	0.3090*	0.0645	-0.0295	0.0651	0.4502*	0.0153	-0.1139*	0.0179
Q14	0.0699#	0.0397	0.0699#	0.0397	0.0482*	0.0105	0.0482*	0.0105
Q15	0.0169	0.0409	0.0169	0.0409	0.0443*	0.0105	0.0443*	0.0105
Q16	0.0214	0.0417	0.0214	0.0417	0.0513*	0.0104	0.0513*	0.0104
Q17	0.1518*	0.0441	-0.0365	0.0457	0.2177*	0.0113	-0.0961*	0.0165
Q18	0.1016*	0.0435	-0.0864#	0.0451	0.2098*	0.0113	-0.1035*	0.0165
Q19	0.1144*	0.0436	-0.0666	0.0453	0.2174*	0.0112	-0.0843*	0.0165
Q20	0.1354*	0.0441	-0.0522	0.0458	0.2379*	0.0113	-0.0748*	0.0166
Q21	0.1119*	0.0442	-0.0678	0.0459	0.2300*	0.0112	-0.0695*	0.0166
Q22	0.1548*	0.0432	-0.0250	0.0448	0.2350*	0.0112	-0.0646*	0.0165
Q23	0.0944*	0.0438	-0.0856#	0.0454	0.2454*	0.0112	-0.0545*	0.0166
Q24	0.1638*	0.0439	-0.0167	0.0456	0.2695*	0.0112	-0.0312#	0.0166
Q25	0.1770*	0.0442	-0.0058	0.0459	0.2858*	0.0113	-0.0188	0.0167
Q26	0.1474*	0.0439	-0.0320	0.0456	0.2908*	0.0112	-0.0082	0.0167
Q27	0.2117*	0.0428	0.0321	0.0446	0.3051*	0.0113	0.0059	0.0168
Q28	0.2352*	0.0414	0.0441	0.0425	0.3370*	0.0103	0.0186	0.0142
Q29	0.2301*	0.0407	0.0374	0.0419	0.3404*	0.0103	0.0192	0.0143
Q30	0.2425*	0.0404	0.0562	0.0416	0.3353*	0.0102	0.0249#	0.0143
Q31	0.2685*	0.0406	0.0771#	0.0418	0.3631*	0.0103	0.0441*	0.0143
Selection bias variables								
λ	-0.2375*	0.0536			-0.3958*	0.0112		
Adjusted R^2		0.1507				0.2999		
Model test		F(50, 6,557) = 24.45; p = 0.0000				F(50, 145,284) = 1,246.11; p = 0.0000		
N		6,608				145,335		

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table A5.7.1. Results of Model 4a: participation selection bias corrected estimates of separate wage equations [A5.7.2] and [A5.7.4] for nurses and all other workers

The results of the decomposition analysis conducted using these results based on equations [5.16] and [5.17] are presented in Table A5.7.2.

	Premium to being a nurse analysed using characteristics of nurses	Premium to being a nurse analysed using characteristics of all other workers
Participation selection bias corrected estimates (Model 4a)		
Differences in variables (= differences in endowments)	0.2875	0.3014
Premium (= differences in returns to endowments)	-0.0018	-0.0157
Differences due to occupation selection bias	0.0791	0.0791
Observed difference in mean lnW	0.3648	0.3648

Table A5.7.2. Results of decomposition analysis for Model 4a

APPENDICES TO CHAPTER 6

Appendix 6.1. Estimating the selection bias variables in the bivariate probit model

```
NAMELIST ; X1 = Z ; X2 = z $

PROCEDURE = Lambda2(W1, W2, P, Nu, Lambda1, Lambda2) $

BIVARIATE ; Lhs = P, Nu

; Rh1 = W1 ; Rh2 = W2 $

CALCULATE ; K1 = Col(W1)

; J1 = K1+1 ; J2 = K1+Col(W2) $

MATRIX ; B1 = Part(B, 1 , K1) ; B2 = Part(B, J1, J2) $

CREATE ; q1 = 2*P-1 ; q2 = 2*Nu-1

; c1 = q1 * b1'W1 ; c2 = q2 * b2'w2

; v1 = (c2 - rho * c1)/sqr(1 - rho*rho) ; v2 = (c1 - rho * c2)/sqr(1 - rho*rho) $

NAMELIST ; V = v1, v2 $

CREATE ; Lambda1 = q1 * N01(c1) * Phi(v1) / Bvn(V,(q1*q2*rho))

; Lambda2 = q2 * N01(c2) * Phi(v2) / Bvn(V, (q1*q2*rho)) $

ENDPROCEDURE

EXECUTE ; PROC = Lambda2(X1, X2, P, Nu, Lambda1, Lambda2) $
```

Appendix 6.2. Estimating the trinomial logit selection model [26]

For nurses (D = 2): [27]

```
LOGIT ; LHS = D ; RHS = z $

(*)MATRIX ; a1 = part(b,1,34) ; a2 = part(b,1,34) ; SG = varb $

(*)CREATE ; p1 = exp(dot(V,a1)) ; p2 = exp(dot(V,a2))

(*); p0 = 1 / (1 + p1 + p2)

(*); p1 = p1 * p0 ; p2 = p2 * p0 $

INCLUDE ; NEW ; D = 2 $

CREATE ; H = inp(p2) ; lambda = n01(H) / phi(H) ; delta = lambda * (H + lambda) $

REGRESS ; lhs = LNWAGE ; rhs = X $

CALCULATE ; thetajsq = b(1) ^ 2

; sigmajsq = sumsqdev / n + thetajsq * xbr(delta)

; rhojsq = thetajsq / sigmajsq $

CREATE ; t = 1 - rhojsq * delta

(*); q1 = -delta*p1*p2; q2 = delta*p2*(1 - p2)$

(*)MATRIX ; F1 = W'[q1]V ; F2= W'[q2]V

(*); F = [F1,F2]

; PSI = sigmajsq * W'[t]W + thetajsq * F * SG * F'

; C = <W'W> * PSI * <W'W>

; Stat(B,C) $
```

[26] In the quadrinomial logit selection model a slightly different set of commands is used. The basic commands are the same with changes to the commands marked with an asterisk (*).
[27] For individuals employed in occupations other than nursing a similar set of commands are used with D = 1

Appendix 6.3. Sample means and standard deviations of variables included in the statistical models

	Entire sample [1]		Nurses [2]		All other workers [2]		Definition
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	
LNWAGE*			1.9471	0.3790	1.5882	0.5125	LN hourly wage
NURSE	0.0359	0.1860					Employed as a nurse=1, 0 otherwise
PART	0.6416	0.4795					Participate in the labour market=1, 0 otherwise
Years of education variables							
YED*	13.2325	2.7559	13.6726	2.1281	13.4141	2.5825	Years of full-time education
YED2*	182.6930	82.2442	191.4690	65.2364	186.6080	78.1909	Years of full-time education squared
Educational attainment variables							
NURSEQUA*	0.0485	0.2149	0.9151	0.2788	0.0244	0.1543	Has a nursing qualification=1, 0 otherwise
PGDEG*	0.0304	0.1717	0.0173	0.1305	0.0411	0.1984	Highest qualification is a postgraduate degree=1, 0 otherwise
DEG*	0.0802	0.2717	0.0878	0.2831	0.1005	0.3007	Highest qualification is a first degree=1, 0 otherwise
ALEVEL*	0.0720	0.2585	0.0090	0.0942	0.0730	0.2602	Highest qualification is A level=1, 0 otherwise
NOQUAL*	0.1983	0.3987	0.0035	0.0588	0.1318	0.3383	Has no qualifications=1, 0 otherwise
Work experience variables							
EXP*			10.3216	8.1480	7.8358	6.9218	Years of experience with current employer
EXP2*			172.9050	242.0870	109.3110	185.9030	Years of experience with current employer squared
Personal characteristic variables							
DISABLE#	0.1568	0.3636	0.0601	0.2377	0.0680	0.2517	Health problems affect paid work =1, 0 otherwise
ETHNIC*	0.0594	0.2364	0.0511	0.2203	0.0383	0.1919	Non-white ethnic group=1, 0 otherwise
NONBRIT*	0.0561	0.2300	0.0618	0.2409	0.0393	0.1944	Non-British nationality=1, 0 otherwise
ETHNBRIT*	0.0191	0.1368	0.0165	0.1273	0.0093	0.0961	Non-white and non-British=1, 0 otherwise
Regional variables							
SEAST*	0.2922	0.4548	0.2641	0.4409	0.2997	0.4581	Lives in the South East of England=1, 0 otherwise
Job characteristic variables							
HOURSPW*			33.6053	10.5120	31.4290	12.9791	Total usual hours worked per week
MANAGE*			0.7437	0.4366	0.2479	0.4318	Employed as a supervisor, manager or foreman=1, 0 otherwise
NWORKERS*			0.8310	0.3748	0.6298	0.4829	25+ workers at workplace=1, 0 otherwise
TEMP*			0.0569	0.2317	0.0754	0.2641	Job is non-permanent or temporary=1, 0 otherwise
Time trend variables							
Q18#	0.0559	0.2298	0.0627	0.2425	0.0554	0.2288	Quarter 18=1, 0 otherwise

Q19	0.0547	0.2273	0.0546	0.2272	0.0555	0.2290	Quarter 19=1, 0 otherwise
Q20	0.0543	0.2266	0.0537	0.2255	0.0541	0.2261	Quarter 20=1, 0 otherwise
Q21	0.0538	0.2256	0.0500	0.2179	0.0548	0.2276	Quarter 21=1, 0 otherwise
Q22	0.0550	0.2279	0.0601	0.2377	0.0558	0.2294	Quarter 22=1, 0 otherwise
Q23	0.0544	0.2268	0.0535	0.2250	0.0555	0.2289	Quarter 23=1, 0 otherwise
Q24	0.0544	0.2269	0.0520	0.2221	0.0557	0.2294	Quarter 24=1, 0 otherwise
Q25	0.0528	0.2237	0.0511	0.2203	0.0534	0.2248	Quarter 25=1, 0 otherwise
Q26	0.0531	0.2242	0.0514	0.2209	0.0543	0.2266	Quarter 26=1, 0 otherwise
Q27	0.0526	0.2233	0.0601	0.2377	0.0534	0.2248	Quarter 27=1, 0 otherwise
Q28[#]	0.1036	0.3047	0.0922	0.2893	0.1013	0.3017	Quarter 28=1, 0 otherwise
Q29	0.1012	0.3017	0.1003	0.3004	0.0985	0.2979	Quarter 29=1, 0 otherwise
Q30	0.1000	0.3000	0.0988	0.2985	0.1004	0.3006	Quarter 30=1, 0 otherwise
Q31	0.0989	0.2986	0.1031	0.3042	0.0969	0.2958	Quarter 31=1, 0 otherwise
Age variables							
AGE	38.6731	11.4875					Years of age
AGE2	1627.5700	900.8340					Years of age squared
Family variables							
PCHILD	0.1324	0.3390					Age 20-29 years and cohabiting or age 25-34 years and married=1, 0 otherwise
COHABIT	0.1000	0.3000					Cohabiting (living as a couple but not married)=1, 0 otherwise
MARRIED	0.5753	0.4943					Married=1, 0 otherwise
Property income variables							
PENSION	0.0186	0.1351					Receives an occupational pension=1, 0 otherwise
NONLABY	121.1280	1059.3200					Non-labour income
N	125,778		3,461		77,233		

[1] Summary statistics for variables included in the participation and occupation selection equations. Includes workers and non-workers.

[2] Summary statistics for variables included in the wage equations. Includes workers only.

* Difference in mean values between nurses and all other workers significant at the 5% level.

[#] Difference in mean values between nurses and all other workers significant at the 10% level.

Table A6.3.1. Sample means and standard deviations

Appendix 6.4. Extended OLS earnings functions for all workers including a dummy variable for whether or not an individual is employed as a nurse

The estimated model is $\ln W_i = \beta X_i + \beta_n Nu_i + U_i$, where Nu is a dummy variable delineating whether or not the individual is employed as a nurse (NURSE = 1, 0 otherwise). This model is estimated using the sub-sample of individuals in the data who participate in the labour market.

	β [1]	Std.Err. [2]
Constant	0.0479	0.0538
NURSE	0.1993*	0.0072
Years of education variables		
YED	0.1225*	0.0072
YED2	-0.0032*	0.0002
Educational attainment variables		
PGDEG	0.3849*	0.0092
DEG	0.2867*	0.0066
ALEVEL	0.0308*	0.0069
NOQUAL	-0.1539*	0.0045
Work experience variables		
EXP	0.0317*	0.0006
EXP2	-0.0006*	0.0000
Personal characteristics variables		
DISABLE	-0.0782	0.0064
ETHNIC	-0.0005	0.0096
NONBRIT	0.0514*	0.0099
ETHNBRIT	-0.1098*	0.0217
Regional variables		
SEAST	0.1603*	0.0035
Job characteristics variables		
HOURSPW	0.0017*	0.0002
MANAGE	0.1441*	0.0038
NWORKERS	0.1308*	0.0034
TEMP	0.0079	0.0071
Time trend variables		
Q18	-0.0082	0.0091
Q19	0.0065	0.0087
Q20	0.0218*	0.0092
Q21	0.0243*	0.0092
Q22	0.0279*	0.0091
Q23	0.0369*	0.0089
Q24	0.0626*	0.0089
Q25	0.0753*	0.0092
Q26	0.0839*	0.0092
Q27	0.0997*	0.0090
Q28	0.1181*	0.0079
Q29	0.1200*	0.0079
Q30	0.1261*	0.0079
Q31	0.1466*	0.0079
Adjusted R^2	0.3048	
Model test	F(32, 80,661) = 1,106.34; p = 0.0000	
N	80,694	

[1] Dependent variable is LNWAGE
[2] Results corrected for heteroscedasticity using White's estimator
* Significant at the 5% level

Table A6.4.1. OLS wage equation estimates with NURSE dummy

Appendix 6.5. Extended earnings functions for all workers with a correction for participation selection bias using the Heckman two-step procedure

This model is estimated using the Heckman two-step procedure. The participation equation is $P_i^* = \delta Z_i + V_i$. The wage equation is $\ln W_i = \beta X_i + \beta_n Nu_i + \beta_{\lambda(p)}\lambda(p)_i + \varepsilon_i$. The participation equation is estimated using the whole sample of individuals in the data (participators and non-participators). The wage equation is estimated using the sub-sample of those who participate.

	δ [1]	Std.Err.
Constant	-2.2896*	0.0897
Age variables		
AGE	0.0739*	0.0025
AGE2	-0.0008*	0.0000
Personal characteristics		
DISABLE	-1.0733*	0.0111
ETHNIC	-0.4529*	0.0197
NONBRIT	-0.3575*	0.0202
ETHNBRIT	-0.0060	0.0395
Family characteristics		
PCHILD	-0.0798*	0.0137
COHABIT	0.2450*	0.0145
MARRIED	0.0540*	0.0106
Property income variables		
PENSION	-0.6746*	0.0298
NONLABY	0.0000*	0.0000
Years of education variables		
YED	0.2012*	0.0100
YED2	-0.0069*	0.0003
Educational attainment variables		
PGDEG	0.5524*	0.0281
DEG	0.3341*	0.0175
ALEVEL	-0.0910*	0.0156
NOQUAL	-0.6186*	0.0107
Regional variables		
SEAST	0.0155#	0.0087
Time trend variables		
Q18	0.0019	0.0231
Q19	0.0483*	0.0233
Q20	0.0039	0.0232
Q21	0.0574*	0.0234
Q22	0.0565*	0.0233
Q23	0.0551*	0.0233
Q24	0.0523*	0.0233
Q25	0.0364	0.0235
Q26	0.0594*	0.0235
Q27	0.0576*	0.0236
Q28	-0.0182	0.0203
Q29	-0.0350#	0.0204
Q30	0.0071	0.0205
Q31	-0.0263	0.0205
Log likelihood function	-70,023.34	
Restricted log likelihood	-82,072.15	
Model test	$\chi^2 = 24{,}097.62$; df = 32; sig. = 0.0000	
N	125,778	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)

* Significant at the 5% level

Significant at the 10% level

Table A6.5.1. Results of participation equation based on entire sample

	β [1]	Std.Err
Constant	0.7550*	0.0477
NURSE	0.1954*	0.0078
Years of education variables		
YED	0.0576*	0.0059
YED2	-0.0009*	0.0002
Educational attainment variables		
PGDEG	0.2189*	0.0124
DEG	0.1825*	0.0083
ALEVEL	0.0745*	0.0077
NOQUAL	0.0305*	0.0082
Years of experience variables		
EXP	0.0269*	0.0007
EXP2	-0.0005*	0.0000
Personal characteristics variables		
DISABLE	0.2930*	0.0134
ETHNIC	0.1445*	0.0115
NONBRIT	0.1540*	0.0110
ETHNBRIT	-0.0851*	0.0219
Regional variables		
SEAST	0.1523*	0.0041
Job characteristics variables		
HOURSPW	0.0018*	0.0001
MANAGE	0.1384*	0.0037
NWORKERS	0.1305*	0.0032
TEMP	0.0109#	0.0059
Time trend variables		
Q18	-0.0097	0.0110
Q19	-0.0090	0.0111
Q20	0.0154	0.0111
Q21	0.0029	0.0112
Q22	0.0070	0.0111
Q23	0.0157	0.0111
Q24	0.0409*	0.0111
Q25	0.0566*	0.0112
Q26	0.0581*	0.0112
Q27	0.0735*	0.0112
Q28	0.1108*	0.0098
Q29	0.1117*	0.0098
Q30	0.1043*	0.0098
Q31	0.1339*	0.0099
Selection bias variables		
$\lambda(p)$	-0.5798*	0.0183
Adjusted R^2	0.3160	
Model test	$F(33, 80{,}660) = 1{,}130.69$; $p = 0.0000$	
N	80,694	

[1] Dependent variable is LNWAGE

* Significant at the 5% level

Significant at the 10% level

Table A6.5.2. Results of wage equation for all workers with NURSE dummy estimated by Heckman two-step procedure with correction for participation selection bias

Appendix 6.6. Extended OLS earnings functions for nurses and other workers

For nurses the estimated model is $\ln W_{ni} = \beta_n X_{ni} + U_{ni}$, which is estimated using data for working nurses only. For all other workers the estimated model is $\ln W_{oi} = \beta_o X_{oi} + U_{oi}$, which is estimated using data for all other workers only.

	Nurses		All other workers	
	β_n [1]	Std.Err. [2]	β_o [1]	Std.Err. [2]
Constant	0.7589*	0.1947	0.0680	0.0536
Years of education variables				
YED	0.1202*	0.0256	0.1178*	0.0072
YED2	-0.0039*	0.0009	-0.0031*	0.0002
Educational attainment variables				
NURSEQUA	0.2233*	0.0281	0.1603*	0.0113
PGDEG	0.2260*	0.0354	0.3865*	0.0095
DEG	0.1069*	0.0187	0.2960*	0.0069
ALEVEL	-0.0670	0.1033	0.0391*	0.0070
NOQUAL	-0.2213	0.1353	-0.1474*	0.0046
Work experience variables				
EXP	0.0213*	0.0022	0.0315*	0.0007
EXP2	-0.0004*	0.0001	-0.0006*	0.0000
Personal characteristic variables				
DISABLE	-0.0531#	0.0301	-0.0809*	0.0065
ETHNIC	0.0037	0.0302	0.0036	0.0100
NONBRIT	0.0603*	0.0249	0.0543*	0.0104
ETHNBRIT	-0.0086	0.0557	-0.1175*	0.0229
Regional variables				
SEAST	0.0680*	0.0147	0.1640*	0.0036
Job characteristic variables				
HOURSPW	-0.0038*	0.0007	0.0018*	0.0002
MANAGE	0.0529*	0.0162	0.1442*	0.0039
NWORKERS	-0.0393*	0.0159	0.1343*	0.0034
TEMP	-0.0776*	0.0347	0.0069	0.0072
Time trend variables				
Q18	-0.0566	0.0400	-0.0072	0.0093
Q19	-0.0370	0.0357	0.0082	0.0089
Q20	-0.0192	0.0340	0.0225*	0.0095
Q21	-0.0334	0.0408	0.0241*	0.0094
Q22	0.0013	0.0354	0.0282*	0.0093
Q23	-0.0561	0.0389	0.0391*	0.0091
Q24	0.0177	0.0341	0.0635*	0.0092
Q25	0.0238	0.0334	0.0763*	0.0095
Q26	-0.0012	0.0429	0.0858*	0.0093
Q27	0.0644#	0.0354	0.1000*	0.0092
Q28	0.0753*	0.0311	0.1189*	0.0081
Q29	0.0739*	0.0316	0.1220*	0.0081
Q30	0.0906*	0.0296	0.1260*	0.0081
Q31	0.1103*	0.0316	0.1474*	0.0081
Adjusted R^2	0.1391		0.3012	
Model test	$F(32, 3,428) = 17.31$; $p = 0.0000$		$F(32, 77,200) = 1041.18$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
[2] Results corrected for heteroscedasticity using White's estimator
* Significant at the 5% level
[#] Significant at the 10% level

Table A6.6.1. OLS wage equations estimated separately for nurses and other workers

Appendix 6.7. Extended earnings functions for nurses and other workers with a correction for participation selection bias using the Heckman two-step procedure

This model is estimated using the Heckman two-step procedure. For individuals employed as nurses the following model is estimated: Participation equation: $P_{ni}^* = \delta Z_{ni} + V_{ni}$; Wage equation: $\ln W_{ni} = \beta_n X_{ni} + \beta_{\lambda(p)n} \lambda(p)_{ni} + \varepsilon_{ni}$. The participation equation is estimated using the sample of participating and non-participating nurses in the data. The wage equation is estimated using the sub-sample of nurses who are working. For individuals employed in occupations other than nursing the following model is estimated: Participation equation: $P_{oi}^* = \delta Z_{oi} + V_{oi}$; Wage equation: $\ln W_{oi} = \beta_o X_{oi} + \beta_{\lambda(p)o} \lambda(p)_{oi} + \varepsilon_{oi}$. The participation equation is estimated using the sample of participating and non-participating non-nurses in the data. The wage equation is estimated using the sub-sample of those who are working and employed in occupations other than nursing.

As explained above we estimate separate participation and wage equations for nurses and other workers. In terms of the participation equation an alternative specification is possible if we are prepared to accept that the participation equations for the two groups (nurses and non-nurses) are the same. In this case the participation equation may be estimated using the whole sample of individuals in the data (all participators and non-participators, nurses and non-nurses). It is possible to test the hypothesis that nurses and all other workers have different participation equations based on the likelihood ratio statistic (see Greene, 2000). The null hypothesis is that the co-efficients of the probit model of participation for nurses and all other workers are the same. The alternative hypothesis is that an altogether different participation equation applies for the two groups of individuals (nurses and all other workers). To test for this we use the probit counterpart to the Chow test. The restricted model in this instance is based on all 125,778 observations in the data. The log-likelihood for the participation equation in this model is -71,711.40. The log-likelihoods based on the 4,511 observations for nurses only and the 121,267 observations for all other workers only are –1,984.716 and –67,812.94, respectively. Therefore the log-likelihood for the unrestricted model with separate equations is the sum, -69,797.656. The χ^2 squared statistic for testing the 31 restrictions of the pooled model is twice the difference between the restricted and unrestricted log-likelihoods, or 3,827.488. The 95% critical value from the χ^2 squared distribution is approximately 44.00. So, at this significance level the null hypothesis that the constant terms and the co-efficients are the same on the probit model of participation for nurses and all other workers is rejected. The conclusion is that it is appropriate to estimate separate participation equations for nurses and all other workers using the methods described.

	Nurses [1]		All other workers [1]	
	δ_n [2]	Std.Err	δ_o [2]	Std.Err
Constant	0.5219	0.6621	2.2175*	0.0907
Age variables				
AGE	0.0655*	0.0213	0.0703*	0.0025
AGE2	-0.0010*	0.0002	-0.0008*	0.00003
Personal characteristic variables				
DISABLE	-1.2404*	0.0621	-1.0677*	0.0112
ETHNIC	0.1293	0.1233	-0.4656*	0.0200
NONBRIT	0.1758	0.1149	-0.3756*	0.0206
ETHNBRIT	-0.2851	0.2389	-0.0021	0.0403
Family variables				
PCHILD	-0.0572	0.0797	-0.0882*	0.0140
COHABIT	0.4716*	0.1225	0.2429*	0.0146
MARRIED	-0.2875*	0.0602	0.0660*	0.0108
Property income variables				
PENSION	-0.9682*	0.1282	-0.6559*	0.0308
NONLABY	-0.0001*	0.00002	-0.00004*	0.000004
Years of education variables				
YED	0.0037	0.0572	0.1986*	0.0101
YED2	-0.0014	0.0018	-0.0068*	0.0003
Educational attainment variables				
PGDEG	0.2600	0.1915	0.5699*	0.0285
DEG	0.1128	0.0877	0.3488*	0.0180
ALEVEL	[3]	[3]	-0.08258	0.0157
NOQUAL	[3]	[3]	-0.6148*	0.0108
Regional variables				
SEAST	-0.1663*	0.0513	0.0216*	0.0089

Time trend variables				
Q18	0.1771	0.1300	-0.0046	0.0235
Q19	0.1605	0.1360	0.0450*	0.0237
Q20	0.2421*	0.1395	-0.0024	0.0236
Q21	0.0945	0.1337	0.0579*	0.0238
Q22	0.0593	0.1289	0.0545*	0.0237
Q23	0.1316	0.1352	0.0515*	0.0237
Q24	0.2103	0.1374	0.0488*	0.0237
Q25	0.0221	0.1313	0.0373	0.0239
Q26	0.1235	0.1349	0.0586*	0.0239
Q27	0.1705	0.1317	0.0521*	0.0240
Q28	0.0028	0.1152	-0.0189	0.0206
Q29	-0.0006	0.1146	-0.0365*	0.0207
Q30	0.1184	0.1163	0.0043	0.0208
Q31	0.0653	0.1144	-0.0288	0.0208
Log likelihood function	-1,984.716		-67812.94	
Restricted log likelihood	-2,447.646		-79452.93	
Model test	$\chi^2 = 925.8598$; df = 30; sig. = 0.0000		$\chi^2 = 23279.99$; df = 32; sig. = 0.0000	
N	4,511		121,267	

[1] Participators and non-participators
[2] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[3] These variables predict PART perfectly and so are excluded
* Significant at the 5% level

Table A6.7.1. Results of separate participation equations for nurses and other workers

	Nurses		All other workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.7956*	0.1865	0.7569*	0.0483
Years of education variables				
YED	0.1172*	0.0241	0.0550*	0.0059
YED2	-0.0037*	0.0008	-0.0008*	0.0002
Educational attainment variables				
NURSEQUA	0.2237*	0.0242	0.1490*	0.0103
PGDEG	0.2222*	0.0467	0.2168*	0.0127
DEG	0.1045*	0.0222	0.1879*	0.0086
ALEVEL	-0.0690	0.0674	0.0792*	0.0077
NOQUAL	-0.2224*	0.1049	0.0327*	0.0082
Work experience variables				
EXP	0.0214*	0.0024	0.0267*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	-0.0233	0.0397	0.2846*	0.0137
ETHNIC	0.0029	0.0337	0.1529*	0.0118
NONBRIT	0.0588*	0.0294	0.1616*	0.0113
ETHNBRIT	-0.0056	0.0641	-0.0935*	0.0226
Regional variables				
SEAST	0.0710*	0.0145	0.1542*	0.0042
Job characteristic variables				
HOURSPW	-0.0039*	0.0006	0.0020*	0.0001
MANAGE	0.0529*	0.0146	0.1389*	0.0038
NWORKERS	-0.0407*	0.0163	0.1340*	0.0033
TEMP	-0.0764*	0.0271	0.0095	0.0060
Time trend variables				
Q18	-0.0599#	0.0350	-0.0069	0.0113
Q19	-0.0407	0.0363	-0.0061	0.0113
Q20	-0.0236	0.0364	0.0179	0.0114
Q21	-0.0356	0.0369	0.0030	0.0114
Q22	-0.0005	0.0352	0.0076	0.0113
Q23	-0.0591	0.0364	0.0190#	0.0113
Q24	0.0136	0.0367	0.0429*	0.0113
Q25	0.0227	0.0367	0.0574*	0.0114
Q26	-0.0039	0.0367	0.0604*	0.0114
Q27	0.0607#	0.0355	0.0754*	0.0115
Q28	0.0744*	0.0322	0.1116*	0.0100
Q29	0.0728*	0.0318	0.1142*	0.0100
Q30	0.0881*	0.0320	0.1051*	0.0100
Q31	0.1087*	0.0317	0.1354*	0.0101
Selection bias variables				
$\lambda(p)$	-0.0443	0.0455	-0.5727*	0.0188
Adjusted R^2	0.1312		0.3122	
Model test	$F(33, 3{,}427) = 16.81$; $p = 0.0000$		$F(33, 77{,}199) = 1{,}063.07$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table A6.7.2. Results of separate wage equations for nurses and all other workers estimated by Heckman two-step procedure with correction for participation selection bias

Appendix 6.8. Extended earnings functions for nurses and other workers with a correction for occupation selection bias using the Heckman two-step procedure

This model is equivalent to Model 5 in Chapter 5. It is estimated using the Heckman two-step procedure. For individuals employed as nurses the following model is estimated: Occupation selection equation: $Nu_i^* = \gamma z_i + v_i$; Wage equation: $\ln W_{ni} = \beta_n X_{ni} + \beta_{\lambda(nu)n}\lambda(nu)_{ni} + \varepsilon_{ni}$. The occupation selection equation is estimated using the sub-sample of individuals in the data who participate. The wage equation is estimated using the sub-sample of nurses who are working. For individuals employed in occupations other than nursing the following model is estimated: Participation equation: $Nu_i^* = \gamma z_i + v_i$; Wage equation: $\ln W_{oi} = \beta_o X_{oi} + \beta_{\lambda(nu)o}\lambda(nu)_{oi} + \varepsilon_{oi}$. The occupation selection equation is estimated using the sub-sample of individuals in the data who choose to participate. The wage equation is estimated using the sub-sample of this group of individuals who are employed in occupations other than nursing.

	γ [1]	Std.Err.
Constant	-2.6536*	0.3938
NURSEQUA	3.0340*	0.0306
Age variables		
AGE	0.0120	0.0102
AGE2	-0.0002#	0.0001
Personal characteristic variables		
DISABLE	-0.2198*	0.0581
ETHNIC	0.2253*	0.0716
NONBRIT	0.3365*	0.0640
ETHNBRIT	-0.0487	0.1387
Family variables		
PCHILD	0.1351*	0.0449
COHABIT	-0.0472	0.0495
MARRIED	-0.1114*	0.0366
Property income variables		
PENSION	-0.0246	0.1386
NONLABY	-0.00003#	0.00002
Years of education variables		
YED	0.0087	0.0451
YED2	-0.0006	0.0015
Educational attainment variables		
PGDEG	-0.7110*	0.0897
DEG	-0.0482	0.0481
ALEVEL	0.0538	0.0684
NOQUAL	-0.3629*	0.0923
Regional variables		
SEAST	-0.0550#	0.0307
Time trend variables		
Q18	0.0032	0.0803
Q19	0.0495	0.0808
Q20	-0.0542	0.0830
Q21	-0.1765*	0.0848
Q22	-0.0020	0.0811
Q23	-0.1145	0.0827
Q24	-0.1083	0.0835
Q25	-0.0759	0.0835
Q26	-0.1519#	0.0840
Q27	0.0407	0.0805
Q28	-0.0483	0.0725
Q29	0.0880	0.0713
Q30	0.0232	0.0719
Q31	0.0812	0.0711
Log likelihood function	-5120.661	
Restricted log likelihood	-14284.75	
Model test	$\chi^2 = 18328.18$; df = 33; sig. = 0.0000	
N	80,694	

[1] Dependent variable is whether the individual is employed as a nurse (NURSE = 1) or not (NURSE = 0)
* Significant at the 5% level
# Significant at the 10% level

Table A6.8.1. Results of occupation selection equation for all workers

	Nurses		All other workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.9557*	0.3592	0.0680#	0.0373
Years of education variables				
YED	0.1174*	0.0243	0.1178*	0.0049
YED2	-0.0038*	0.0008	-0.0031*	0.0002
Educational attainment variables				
NURSEQUA	0.0815	0.2240	0.1510*	0.0598
PGDEG	0.2549*	0.0650	0.3868*	0.0091
DEG	0.1082*	0.0221	0.2961*	0.0063
ALEVEL	-0.0714	0.0675	0.0391*	0.0062
NOQUAL	-0.1993#	0.1101	-0.1473*	0.0050
Work experience variables				
EXP	0.0214*	0.0024	0.0315*	0.0007
EXP2	-0.0004*	0.0001	-0.0006*	0.0000
Personal characteristic variables				
DISABLE	-0.0446	0.0287	-0.0808*	0.0062
ETHNIC	-0.0031	0.0354	0.0035	0.0093
NONBRIT	0.0489	0.0345	0.0541*	0.0093
ETHNBRIT	-0.0084	0.0641	-0.1175*	0.0205
Regional variables				
SEAST	0.0704*	0.0147	0.1641*	0.0035
Job characteristic variables				
HOURSPW	-0.0038*	0.0006	0.0018*	0.0001
MANAGE	0.0529*	0.0146	0.1442*	0.0038
NWORKERS	-0.0407*	0.0164	0.1342*	0.0033
TEMP	-0.0776*	0.0271	0.0069	0.0061
Time trend variables				
Q18	-0.0570	0.0349	-0.0072	0.0093
Q19	-0.0389	0.0362	0.0082	0.0093
Q20	-0.0174	0.0363	0.0225*	0.0093
Q21	-0.0271	0.0383	0.0242*	0.0093
Q22	0.0011	0.0353	0.0282*	0.0093
Q23	-0.0519	0.0369	0.0391*	0.0093
Q24	0.0216	0.0371	0.0635*	0.0093
Q25	0.0269	0.0371	0.0763*	0.0094
Q26	0.0044	0.0377	0.0858*	0.0094
Q27	0.0632#	0.0354	0.1000*	0.0094
Q28	0.0772*	0.0324	0.1189*	0.0082
Q29	0.0710*	0.0321	0.1220*	0.0083
Q30	0.0903*	0.0319	0.1260*	0.0082
Q31	0.1081*	0.0319	0.1474*	0.0083
Selection bias variables				
λ(nu)	-0.0603	0.0948	0.0095	0.0605
Adjusted R^2	0.1309		0.3122	
Model test	$F(33, 3{,}427) = 16.79$; $p = 0.0000$		$F(33, 77{,}199) = 1{,}009.62$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table A6.8.2. Results of separate wage equations for nurses and all other workers estimated by Heckman two-step procedure with correction for occupation selection bias

Appendix 6.9. Extended earnings functions for nurses and other workers with a correction for occupation selection bias using the Heckman two-step procedure (version 2)

This model is similar to version 1 in Appendix 6.7. The only difference is in the estimation procedure in that the occupation selection equation is estimated on the entire sample in the data (n = 125,778) – in version 1 the occupation selection is estimated using data for workers only (n = 80,694). This means that in this version the nurse/not-nurse occupation selection decision is estimated on both participating and non-participating nurses and other workers, as defined in the main text. The wage equations are estimated as before.

	γ [1]	Std.Err.
Constant	-2.6667*	0.3002
NURSEQUA	3.3154*	0.0289
Age variables		
AGE	-0.0029	0.0093
AGE2	-0.000003	0.0001
Personal characteristic variables		
DISABLE	0.0234	0.0400
ETHNIC	0.1525*	0.0650
NONBRIT	0.2737*	0.0593
ETHNBRIT	-0.0964	0.1213
Family variables		
PCHILD	0.1446*	0.0426
COHABIT	-0.0337	0.0495
MARRIED	-0.0705*	0.0341
Property income variables		
PENSION	0.3376*	0.0833
NONLABY	-0.00002#	0.00002
Years of education variables		
YED	0.0127	0.0304
YED2	-0.0006	0.0010
Educational attainment variables		
PGDEG	-0.7726*	0.0871
DEG	-0.0353	0.0465
ALEVEL	0.0307	0.0653
NOQUAL	-0.5561*	0.0863
Regional variables		
SEAST	-0.0355	0.0288
Time trend variables		
Q18	-0.0249	0.0754
Q19	0.0286	0.0760
Q20	-0.0961	0.0785
Q21	-0.1808*	0.0791
Q22	-0.0117	0.0760
Q23	-0.1403#	0.0778
Q24	-0.1384#	0.0788
Q25	-0.0817	0.0777
Q26	-0.1666*	0.0788
Q27	0.0105	0.0758
Q28	-0.0515	0.0673
Q29	0.0734	0.0665
Q30	0.0011	0.0674
Q31	0.0722	0.0663
Log likelihood function	-5,694.127	
Restricted log likelihood	-19,441.73	
Model test	$\chi^2 = 27{,}495.20$; df = 33; sig. = 0.0000	
N	125,778	

[1] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0)

* Significant at the 5% level

[#] Significant at the 10% level

Table A6.9.1. Results of occupation selection equation for the entire sample

	Nurses		All other workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.8470#	0.4735	0.0681#	0.0373
Years of education variables				
YED	0.1195*	0.0241	0.1178*	0.0049
YED2	-0.0038*	0.0008	-0.0031*	0.0002
Educational attainment variables				
NURSEQUA	0.1534	0.3473	0.2268*	0.0708
PGDEG	0.2396*	0.0821	0.3845*	0.0091
DEG	0.1074*	0.0221	0.2958*	0.0063
ALEVEL	-0.0681	0.0675	0.0390*	0.0062
NOQUAL	-0.2070	0.1266	-0.1478*	0.0050
Work experience variables				
EXP	0.0213*	0.0024	0.0315*	0.0007
EXP2	-0.0004*	0.0001	-0.0006*	0.00002
Personal characteristic variables				
DISABLE	-0.0533*	0.0253	-0.0809*	0.0061
ETHNIC	0.0017	0.0350	0.0040	0.0093
NONBRIT	0.0562	0.0358	0.0551*	0.0092
ETHNBRIT	-0.0076	0.0642	-0.1177*	0.0205
Regional variables				
SEAST	0.0686*	0.0145	0.1639*	0.0034
Job characteristic variables				
HOURSPW	-0.0038*	0.0006	0.0018*	0.0001
MANAGE	0.0529*	0.0146	0.1442*	0.0038
NWORKERS	-0.0396*	0.0163	0.1343*	0.0033
TEMP	-0.0777*	0.0271	0.0069	0.0061
Time trend variables				
Q18	-0.0564	0.0348	-0.0072	0.0093
Q19	-0.0374	0.0361	0.0083	0.0093
Q20	-0.0178	0.0368	0.0223*	0.0093
Q21	-0.0307	0.0393	0.0237*	0.0093
Q22	0.0014	0.0352	0.0281*	0.0093
Q23	-0.0538	0.0379	0.0388*	0.0093
Q24	0.0198	0.0380	0.0632*	0.0093
Q25	0.0252	0.0373	0.0761*	0.0094
Q26	0.0014	0.0388	0.0854*	0.0094
Q27	0.0645#	0.0353	0.1000*	0.0094
Q28	0.0762*	0.0325	0.1187*	0.0082
Q29	0.0730*	0.0321	0.1222*	0.0083
Q30	0.0908*	0.0319	0.1260*	0.0082
Q31	0.1095*	0.0319	0.1476*	0.0083
Selection bias variables				
λ(nu)	-0.0273	0.1355	-0.0602	0.0634
Adjusted R^2	0.1308		0.3012	
Model test	$F(33, 3{,}427) = 16.78$; $p = 0.0000$		$F(33, 77{,}199) = 1{,}009.66$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
\# Significant at the 10% level

Table A6.9.2. Results of separate wage equations for nurses and all other workers estimated by Heckman two-step procedure with correction for occupation selection bias

Appendix 6.10. Full results of the statistical models

The tables below show the full set of results for the statistical models as discussed in the text (Section 6.5), including the time trend variables.

A6.10.1. Bivariate probit selection model

	Participation equation		Occupation selection equation	
	δ [1]	Std.Err.	γ [2]	Std.Err.
Constant	-2.2921*	0.0749	-2.6670*	0.3572
NURSEQUA			3.3397*	0.0323
Age variables				
AGE	0.0739*	0.0025	-0.0018	0.0101
AGE2	-0.0008*	0.00003	-0.00002	0.0001
Personal characteristics				
DISABLE	-1.0734*	0.0111	0.0312	0.0556
ETHNIC	-0.4530*	0.0199	0.1673*	0.0648
NONBRIT	-0.3575*	0.0201	0.2883*	0.0532
ETHNBRIT	-0.0055	0.0391	-0.1100	0.1184
Family variables				
PCHILD	-0.0799*	0.0136	0.1487*	0.0435
COHABIT	0.2449*	0.0146	-0.0368	0.0503
MARRIED	0.0541*	0.0108	-0.0807*	0.0341
Property income variables				
PENSION	-0.6753*	0.0298	0.3288*	0.1105
NONLABY	-0.00004*	0.000002	-0.00003	0.00002
Years of education variables				
YED	0.2015*	0.0072	0.0116	0.0344
YED2	-0.0069*	0.0002	-0.0006	0.0011
Educational attainment variables				
PGDEG	0.5524*	0.0271	-0.7776*	0.0706
DEG	0.3340*	0.0171	-0.0415	0.0401
ALEVEL	-0.0913*	0.0155	0.0312	0.0639
NOQUAL	-0.6185*	0.0108	-0.5505*	0.0887
Regional variables				
SEAST	0.0155#	0.0086	-0.0372	0.0286
Time trend variables				
Q18	0.0020	0.0231	-0.0194	0.0817
Q19	0.0483*	0.0233	0.0316	0.0803
Q20	0.0038	0.0233	-0.0901	0.0853
Q21	0.0574*	0.0235	-0.1791*	0.0877
Q22	0.0565*	0.0233	-0.0115	0.0842
Q23	0.0552*	0.0234	-0.1339	0.0820
Q24	0.0523*	0.0233	-0.1328	0.0854
Q25	0.0364	0.0235	-0.0804	0.0842
Q26	0.0596*	0.0235	-0.1652	0.0846

Q27	0.0575*	0.0235	0.0144	0.0818
Q28	-0.0180	0.0203	-0.0475	0.0725
Q29	-0.0351#	0.0204	0.0769	0.0715
Q30	0.0071	0.0205	0.0058	0.0724
Q31	-0.0262	0.0205	0.0758	0.0709
ρ_v		-0.1386*		
Log likelihood function		-75,693.25		
N		125,778		

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[2] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0)
* Significant at the 5% level
# Significant at the 10% level

Table A6.10.1 (based on Table 6.3 in the main text). Results of participation and occupation selection equations estimated jointly by bivariate probit

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.8564*	0.2084	0.2250*	0.0398
Years of education variables				
YED	0.1052*	0.0247	0.0997*	0.0052
YED2	-0.0033*	0.0008	-0.0025*	0.0002
Educational attainment variables				
NURSEQUA	0.2609*	0.1041	0.1208#	0.0673
PGDEG	0.1863*	0.0552	0.3661*	0.0092
DEG	0.08718*	0.0241	0.2826*	0.0064
ALEVEL	-0.0584	0.0677	0.0467*	0.0063
NOQUAL	-0.0844	0.1461	-0.1123*	0.0059
Work experience variables				
EXP	0.0208*	0.0024	0.0308*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristics variables				
DISABLE	0.0238	0.0505	-0.0112	0.0088
ETHNIC	0.0324	0.0376	0.0259*	0.0095
NONBRIT	0.0843*	0.0336	0.0683*	0.0093
ETHNBRIT	0.0029	0.0644	-0.1075*	0.0205
Regional variables				
SEAST	0.0653*	0.0143	0.1632*	0.0034
Job characteristics variables				
HOURSPW	-0.0037*	0.0006	0.0018*	0.0001
MANAGE	0.0526*	0.0147	0.1435*	0.0038
NWORKERS	-0.0377*	0.0164	0.1341*	0.0033
TEMP	-0.0749*	0.0273	0.0073	0.0061
Time trend variables				
Q18	-0.0563	0.0349	-0.0074	0.0093
Q19	-0.0411	0.0362	0.0069	0.0093
Q20	-0.0212	0.0364	0.0221*	0.0093
Q21	-0.0402	0.0374	0.0221*	0.0093
Q22	-0.0019	0.0353	0.0254*	0.0093
Q23	-0.0608#	0.0367	0.0367*	0.0093
Q24	0.0121	0.0370	0.0611*	0.0093
Q25	0.0190	0.0370	0.0741*	0.0094

Q26	-0.0086	0.0373	0.0832*	0.0094
Q27	0.0585#	0.0356	0.0964*	0.0094
Q28	0.0742*	0.0323	0.11828	0.0082
Q29	0.0725*	0.0319	0.1206*	0.0083
Q30	0.0861*	0.0321	0.1230*	0.0082
Q31	0.1101*	0.0318	0.1453*	0.0083
Selection bias variables				
λ(p)	-0.1193	0.0746	-0.0532*	0.0047
λ(nu)	0.0285	0.0419	0.0246	0.0555
Adjusted R^2	0.1322		0.3023	
Model test	$F(34, 3{,}426) = 16.50$; $p = 0.0000$		$F(34, 77{,}198) = 985.22$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table A6.10.2 (based on Table 6.4 in the main text). Results of selection-bias corrected wage equation estimated in the bivariate probit selection model

A6.10.2. Bivariate probit selection model with censoring

	Participation equation		Occupation selection equation	
	δ [1]	Std.Err.	γ [2]	Std.Err.
Constant	-2.2896*	0.0897	-2.6536*	0.3938
NURSEQUA			3.0340*	0.0306
Age variables				
AGE	0.0739*	0.0025	0.0120	0.0102
AGE2	-0.0008*	0.0000	-0.0002#	0.0001
Personal characteristics variables				
DISABLE	-1.0733*	0.0111	-0.2198*	0.0581
ETHNIC	-0.4529*	0.0197	0.2253*	0.0716
NONBRIT	-0.3575*	0.0202	0.3365*	0.0640
ETHNBRIT	-0.0060	0.0395	-0.0487	0.1387
Family variables				
PCHILD	-0.0798*	0.0137	0.1351*	0.0449
COHABIT	0.2450*	0.0145	-0.0472	0.0495
MARRIED	0.0540*	0.0106	-0.1114*	0.0366
Property income variables				
PENSION	-0.6746*	0.0298	-0.0246	0.1386
NONLABY	-0.00004*	0.000005	-0.00003#	0.00002
Years of education				
YED	0.2012*	0.0100	0.0087	0.0451
YED2	-0.0069*	0.0003	-0.0006	0.0015
Educational attainment variables				
PGDEG	0.5524*	0.0281	-0.7110*	0.0897
DEG	0.3341*	0.0175	-0.0482	0.0481
ALEVEL	-0.0910*	0.0156	0.0538	0.0684
NOQUAL	-0.6186*	0.0107	-0.3629*	0.0923
Regional variables				
SEAST	0.0155#	0.0087	-0.0550#	0.0307

Time trend variables				
Q18	0.0019	0.0231	0.0032	0.0803
Q19	0.0483*	0.0233	0.0495	0.0808
Q20	0.0039	0.0232	-0.0542	0.0830
Q21	0.0574*	0.0234	-0.1765*	0.0848
Q22	0.0565*	0.0233	-0.0020	0.0811
Q23	0.0551*	0.0233	-0.1145	0.0827
Q24	0.0523*	0.0233	-0.1083	0.0835
Q25	0.0364	0.0235	-0.0759	0.0835
Q26	0.0594*	0.0235	-0.1519#	0.0840
Q27	0.0576*	0.0236	0.0407	0.0805
Q28	-0.0182	0.0203	-0.0483	0.0725
Q29	-0.0350#	0.0204	0.0880	0.0713
Q30	0.0071	0.0205	0.0232	0.0719
Q31	-0.0263	0.0205	0.0812	0.0711
Log likelihood function	-70,023.34		-5,120.661	
Restricted log likelihood	-82,072.15		-14,284.75	
Model test	$\chi^2 = 24{,}097.62$; df = 32; sig. = 0.0000		$\chi^2 = 18{,}328.18$; df = 33; sig. = 0.0000	
N	125,778		80,694	

Significant at the 10% level

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[2] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0). This only observed when PART = 1.
* Significant at the 5% level

Table A6.10.3 (based on Table 6.5 in the main text). Results of participation and occupation selection equations estimated by independent probit with censoring

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	1.5009*	0.3973	0.7690*	0.0420
Years of education variables				
YED	0.0886*	0.0259	0.0532*	0.0052
YED2	-0.0027*	0.0009	-0.0007*	0.0002
Educational attainment variables				
NURSEQUA	-0.1074	0.2319	0.0449	0.0595
PGDEG	0.2239*	0.0659	0.2246*	0.0101
DEG	0.0671*	0.0255	0.1923*	0.0069
ALEVEL	-0.0641	0.0680	0.0817*	0.0063
NOQUAL	-0.0790	0.1169	0.0358*	0.0071
Work experience variables				
EXP	0.0204*	0.0024	0.0267*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	0.1396*	0.0636	0.2883*	0.0121
ETHNIC	0.0488	0.0389	0.1468*	0.0101
NONBRIT	0.0858*	0.0363	0.1533*	0.0096
ETHNBRIT	0.0101	0.0644	-0.0935*	0.0203
Regional variables				
SEAST	0.0699*	0.0147	0.1563*	0.0034

Job characteristic variables				
HOURSPW	-0.0038*	0.0006	0.0019*	0.0001
MANAGE	0.0516*	0.0147	0.1386*	0.0038
NWORKERS	-0.0375*	0.0165	0.1337*	0.0033
TEMP	-0.0755*	0.0272	0.0100#	0.0060
Time trend variables				
Q18	-0.0572	0.0349	-0.0087	0.0092
Q19	-0.0517	0.0365	-0.0071	0.0092
Q20	-0.0191	0.0363	0.0165#	0.0093
Q21	-0.0314	0.0383	0.0040	0.0093
Q22	-0.0069	0.0354	0.0072	0.0092
Q23	-0.0573	0.0370	0.0188*	0.0092
Q24	0.0153	0.0371	0.0426*	0.0092
Q25	0.0210	0.0371	0.0582*	0.0093
Q26	-0.0023	0.0378	0.0611*	0.0093
Q27	0.0474	0.0358	0.0741*	0.0093
Q28	0.0770*	0.0324	0.1119*	0.0081
Q29	0.0618#	0.0323	0.1135*	0.0082
Q30	0.0790*	0.0322	0.1044*	0.0082
Q31	0.0989*	0.0321	0.1345*	0.0083
Selection bias variables				
$\lambda(p)$	-0.2811*	0.0866	-0.5741*	0.0162
$\lambda(nu)$	-0.1371	0.0979	0.1072#	0.0601
Adjusted R^2	0.1333		0.3124	
Model test	F(34, 3,426) = 16.66; p = 0.0000		F(34, 77,198) = 1,032.81; p = 0.0000	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
# Significant at the 10% level

Table A6.10.4 (based on Table 6.6 in the main text). Results of selection-bias corrected wage equation estimated in the independent probit selection model with censoring

A6.10.3. Trinomial logit selection model

	Employed as a nurse (D = 2)		All other workers (D = 1)	
	ψ_2 [1]	Std.Err.	ψ_1 [1]	Std.Err.
Constant	-8.5936*	0.8155	-3.9246*	0.1689
NURSEQUA	5.9305*	0.0787	-0.2499*	0.0430
Age variables				
AGE	0.1499*	0.0206	0.1159*	0.0042
AGE2	-0.0020*	0.0002	-0.0013*	0.0001
Personal characteristic variables				
DISABLE	-2.0736*	0.0882	-1.7600*	0.0188
ETHNIC	-0.1822	0.1372	-0.7669*	0.0328
NONBRIT	0.0600	0.1276	-0.6175*	0.0340
ETHNBRIT	-0.0321	0.2739	-0.0082	0.0665
Family variables				
PCHILD	0.0918	0.0851	-0.1521*	0.0233
COHABIT	0.3987*	0.0990	0.4138*	0.0247

MARRIED	-0.2052*	0.0658	0.1041*	0.0180
Property income variables				
PENSION	-1.4514*	0.1982	-1.1268*	0.0506
NONLABY	-0.0001*	0.00003	-0.0001*	0.000007
Years of education variables				
YED	0.2824*	0.0942	0.3621*	0.0196

Significant at the 10% level

YED2	-0.0111*	0.0031	-0.0125*	0.0006
Educational attainment variables				
PGDEG	-0.2655	0.1644	1.0548*	0.0518
DEG	0.3855*	0.0924	0.6239*	0.0312
ALEVEL	0.0860	0.1940	-0.1085*	0.0263
NOQUAL	-2.2137*	0.2974	-0.9732*	0.0179
Regional variables				
SEAST	-0.1369*	0.0568	0.0395*	0.0149
Time trend variables				
Q18	0.0788	0.1453	0.0025	0.0390
Q19	0.1625	0.1506	0.0822*	0.0394
Q20	0.0267	0.1502	0.0106	0.0393
Q21	-0.1246	0.1494	0.0979*	0.0397
Q22	0.1011	0.1467	0.0932*	0.0395
Q23	-0.0501	0.1487	0.0968*	0.0396
Q24	0.0126	0.1496	0.0939*	0.0395
Q25	-0.0608	0.1497	0.0638	0.0398
Q26	-0.1075	0.1492	0.1101*	0.0399
Q27	0.2028	0.1476	0.0939*	0.0400
Q28	-0.1298	0.1313	-0.0248	0.0342
Q29	0.0345	0.1315	-0.0601#	0.0345
Q30	0.0612	0.1317	0.0110	0.0347
Q31	0.0217	0.1310	-0.0475	0.0346
Log likelihood function		-74,819.60		
Restricted log likelihood		-96,356.90		
Model test		$\chi^2 = 43{,}074.62$; df = 66; sig. = 0.0000		
N		125,778		

[1] The reference group is non-participators (D = 0)
* Significant at the 5% level

Table A6.10.5 (based on Table 6.7 in the main text). Results of trinomial logit selection model

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	1.1972*	0.2578	1.0356*	0.0772
Years of education variables				
YED	0.1020*	0.0252	0.0682*	0.0069
YED2	-0.0032*	0.0008	-0.0013*	0.0002
Educational attainment variables				
NURSEQUA	0.2298*	0.0247	0.1750*	0.0117
PGDEG	0.1858*	0.0505	0.3029*	0.0179
DEG	0.0814*	0.0250	0.2430*	0.0104
ALEVEL	-0.0617	0.0681	0.0601*	0.0157

NOQUAL	-0.1667	0.1063	-0.0514*	0.0199
Work experience variables				
EXP	0.0206*	0.0024	0.0290*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.0000
Personal characteristic variables				
DISABLE	0.0712	0.0569	0.1582*	0.0151
ETHNIC	0.0433	0.0378	0.0867*	0.0133
NONBRIT	0.0933*	0.0328	0.1127*	0.0128
ETHNBRIT	0.0110	0.0647	-0.0798*	0.0242
Regional variables				
SEAST	0.0650*	0.0145	0.1591*	0.0056
Job characteristic variables				
HOURSPW	-0.0038*	0.0006	0.0019*	0.0001
MANAGE	0.0522*	0.0146	0.1417*	0.0038
NWORKERS	-0.0363*	0.0163	0.1339*	0.0033
TEMP	-0.0759*	0.0271	0.0092	0.0061
Time trend variables				
Q18	-0.0564	0.0356	-0.0080	0.0144
Q19	-0.0433	0.0370	0.0006	0.0147
Q20	-0.0220	0.0370	0.0186	0.0147
Q21	-0.0408	0.0378	0.0137	0.0147
Q22	-0.0036	0.0361	0.0180	0.0145
Q23	-0.0626#	0.0371	0.0280#	0.0147
Q24	0.0105	0.0374	0.0519*	0.0147
Q25	0.0183	0.0376	0.0667*	0.0148
Q26	-0.0104	0.0376	0.0719*	0.0148
Q27	0.0559	0.0363	0.0874*	0.0147
Q28	0.0736*	0.0329	0.1149*	0.0129
Q29	0.0701*	0.0325	0.1179*	0.0129
Q30	0.0836*	0.0327	0.1148*	0.0130
Q31	0.1065*	0.0324	0.1409*	0.0129
Selection bias variables				
λ	-0.3384*	0.1384	-0.6654*	0.0391
Adjusted R^2	0.1323		0.3016	
Model test	F(33, 3,427) = 16.98; p = 0.0000		F(33, 77,199) = 1,033.59; p = 0.0000	
N	3,461		77,233	

[1] Dependent variable is LNWAGE
* Significant at the 5% level
\# Significant at the 10% level

Table A6.10.6 (based on Table 6.8 in the main text). Results of selection-bias corrected wage equation estimated in the trinomial logit selection model

A6.10.4. Quadrinomial logit selection model

	Participating nurses (D = 3)		Participating non-nurses (D = 2)		Non-participating nurses (D = 1)	
	ψ_3 [1]	Std.Err.	ψ_2 [1]	Std.Err.	ψ_1 [1]	Std.Err.
Constant	-25.9419*	0.8338	-4.0275*	0.1709	-16.8345*	1.1886

NURSEQUA [2]						
Age variables						
AGE	0.3521*	0.0151	0.1170*	0.0043	0.2093*	0.0280
AGE2	-0.0039*	0.0002	-0.0013*	0.0001	-0.0017*	0.0003
Personal characteristic variables						
DISABLE	-1.7860*	0.0739	-1.7547*	0.0189	0.2604*	0.0686
	# Significant at the 10% level					
ETHNIC	-0.5703*	0.0995	-0.7752*	0.0330	-0.4307*	0.1673
NONBRIT	-0.2438*	0.0893	-0.6271*	0.0341	-0.3620*	0.1670
ETHNBRIT	-0.1185	0.1896	-0.0057	0.0668	-0.0166	0.3312
Family variables						
PCHILD	0.1588*	0.0614	-0.1517*	0.0235	0.3522*	0.1240
COHABIT	0.2683*	0.0695	0.4094*	0.0248	-0.5280*	0.2056
MARRIED	-0.0924#	0.0486	0.1120*	0.0181	0.2582*	0.0811
Property income variables						
PENSION	-1.0898*	0.1743	-1.0850*	0.0513	0.5240*	0.1126
NONLABY	-0.0001*	0.00002	-0.0001*	0.000007	0.000008	0.00002
Years of education variables						
YED	2.2810*	0.1093	0.3707*	0.0199	1.0237*	0.1450
YED2	-0.0732*	0.0037	-0.0126*	0.0006	-0.0298*	0.0049
Educational attainment variables						
PGDEG	-0.4170*	0.1454	1.0145*	0.0521	-1.1219*	0.3044
DEG	-0.0994	0.0755	0.6082*	0.0314	-0.4126*	0.1418
ALEVEL	-2.7050*	0.1827	-0.1383*	0.0263	-34.4990	2927653
NOQUAL	-4.5638*	0.2905	-1.0069*	0.0179	-35.3505	1927307
Regional variables						
SEAST	-0.2606*	0.0422	0.0413*	0.0150	0.0352	0.0715
Time trend variables						
Q18	0.0855	0.1056	-0.0050	0.0393	-0.2595	0.1821
Q19	0.0315	0.1090	0.0739#	0.0397	-0.2777	0.1885
Q20	-0.0479	0.1093	-0.0040	0.0396	-0.5447*	0.1980
Q21	-0.0359	0.1113	0.0939*	0.0400	-0.1016	0.1793
Q22	0.1253	0.1068	0.0910*	0.0399	-0.0785	0.1778
Q23	0.0220	0.1095	0.0867*	0.0399	-0.3333#	0.1887
Q24	-0.0417	0.1102	0.0821*	0.0398	-0.4250*	0.1930
Q25	-0.0497	0.1107	0.0603	0.0401	-0.1062	0.1778
Q26	-0.0055	0.1106	0.0994*	0.0402	-0.2927	0.1869
Q27	0.1438	0.1069	0.0863*	0.0403	-0.2795	0.1842
Q28	-0.1803#	0.0967	-0.0317	0.0345	-0.2151	0.1556
Q29	-0.0895	0.0956	-0.0650#	0.0347	-0.1775	0.1584
Q30	-0.0720	0.0959	0.0032	0.0349	-0.2963#	0.1631
Q31	-0.0478	0.0953	-0.0511	0.0349	-0.1043	0.1569
Log likelihood function			-87,120.56			
Restricted log likelihood			-101,342.3			
Model test			$\chi^2 = 28,443.49$; df = 96; sig. = 0.0000			
N			125,778			

[1] The reference group is non-participating non-nurses (D = 0)
[2] NURSEQUA predicts D = 0 and D = 1 perfectly and so is omitted
* Significant at the 5% level

Table A6.10.7 (based on Table 6.9 in the main text). Results of quadrinomial logit selection model

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.8921*	0.2129	0.4221*	0.0383
Years of education variables				
YED	0.1127*	0.0247	0.1275*	0.0049
YED2	-0.0036*	0.0008	-0.0034*	0.0002
Educational attainment variables				
NURSEQUA	0.2200*	0.0245	0.1351*	0.0100
PGDEG	0.2368*	0.0475	0.4188*	0.0089
DEG	0.1113*	0.0224	0.2969*	0.0062
ALEVEL	0.1510	0.1894	0.1360*	0.0310
NOQUAL	-0.0010	0.2075	0.9725	0.3014
Work experience variables				
EXP	0.0208*	0.0024	0.0264*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.00002
Personal characteristic variables				
DISABLE	-0.0571*	0.0256	-0.0992*	0.0061
ETHNIC	0.0052	0.0338	0.0234*	0.0093
NONBRIT	0.0636*	0.0296	0.0684*	0.0091
ETHNBRIT	-0.0053	0.0643	-0.1389*	0.0203
Regional variables				
SEAST	0.0668*	0.0143	0.1569*	0.0034
Job characteristic variables				
HOURSPW	-0.0037*	0.0006	0.0023*	0.0001
MANAGE	0.0525*	0.0147	0.1402*	0.0038
NWORKERS	-0.0361*	0.0165	0.1367*	0.0033
TEMP	-0.0785*	0.0273	-0.0011	0.0060
Time trend variables				
Q18	-0.0526	0.0351	0.0053	0.0092
Q19	-0.0333	0.0363	0.0194*	0.0092
Q20	-0.0128	0.0366	0.0473*	0.0093
Q21	-0.0327	0.0370	0.0240*	0.0092
Q22	0.0029	0.0353	0.0248*	0.0092
Q23	-0.0518	0.0365	0.0495*	0.0092
Q24	0.0227	0.0368	0.0783*	0.0092
Q25	0.0245	0.0368	0.0722*	0.0093
Q26	0.0012	0.0368	0.0921*	0.0093
Q27	0.0673#	0.0355	0.1024*	0.0093
Q28	0.0779*	0.0324	0.1178*	0.0081
Q29	0.0748*	0.0319	0.1171*	0.0082
Q30	0.0926*	0.0320	0.1273*	0.0082
Q31	0.1098*	0.0318	0.1372*	0.0082
Selection bias variables				
λ	-0.0352	0.0286	-0.1786*	0.0049
Adjusted R^2	0.1312		0.3128	
Model test	$F(33, 3,427) = 16.83$; $p = 0.0000$		$F(33, 77,199) = 1,066.31$; $p = 0.0000$	
N	3,461		77,233	

[1] Dependent variable is LNWAGE

* Significant at the 5% level

\# Significant at the 10% level

Table A6.10.8 (based on Table 6.10 in the main text). Results of selection-bias corrected wage equation estimated in the quadrinomial logit selection model

Appendix 6.11. Extended earnings functions for all workers with a correction for participation and occupation selection bias using the independent probit model

This model is identical to the bivariate probit model with one important difference: in the independent probit model equations [6.1] - [6.2] (the participation equations and occupation selection equations) are estimated separately as independent probits. The selection bias correction terms are then computed and included in separate wage equations for nurses and other workers.

	Participation equation		Occupation selection equation	
	$\delta^{1,2}$	Std.Err.	$\gamma^{2,3}$	Std.Err.
Constant	-2.2896*	0.0897	-2.6667*	0.3002
NURSEQUA			3.3154*	0.0289
Age variables				
AGE	0.0739*	0.0025	-0.0029	0.0093
AGE2	-0.0008*	0.00003	-0.000003	0.0001
Personal characteristic variables				
DISABLE	-1.0733*	0.0111	0.0234	0.0400
ETHNIC	-0.4529*	0.0197	0.1525*	0.0650
NONBRIT	-0.3575*	0.0202	0.2737*	0.0593
ETHNBRIT	-0.0060	0.0395	-0.0964	0.1213
Family variables				
PCHILD	-0.0798*	0.0137	0.1446*	0.0426
COHABIT	0.2450*	0.0145	-0.0337	0.0495
MARRIED	0.0540*	0.0106	-0.0705*	0.0341
Property income variables				
PENSION	-0.6746*	0.0298	0.3376*	0.0833
NONLABY	-0.00004*	0.000003	-0.00002#	0.00002
Years of education variables				
YED	0.2012*	0.0100	0.0127	0.0304
YED2	-0.0069*	0.0003	-0.0006	0.0010
Educational attainment variables				
PGDEG	0.5524*	0.0281	-0.7726*	0.0871
DEG	0.3341*	0.0175	-0.0353	0.0465
ALEVEL	-0.0910*	0.0156	0.0307	0.0653
NOQUAL	-0.6186*	0.0107	-0.5561*	0.0863
Regional variables				
SEAST	0.0155#	0.0087	-0.0355	0.0288
Time trend variables				
Q18	0.0019	0.0231	-0.0249	0.0754
Q19	0.0483*	0.0233	0.0286	0.0760
Q20	0.0039	0.0232	-0.0961	0.0785
Q21	0.0574*	0.0234	-0.1808*	0.0791
Q22	0.0565*	0.0233	-0.0117	0.0760

Q23	0.0551*	0.0233	-0.1403#	0.0778
Q24	0.0523*	0.0233	-0.1384#	0.0788
Q25	0.0364	0.0235	-0.0817	0.0777
Q26	0.0594*	0.0235	-0.1666*	0.0788
Q27	0.0576*	0.0236	0.0105	0.0758
Q28	-0.0182	0.0203	-0.0515	0.0673
Q29	-0.0350#	0.0204	0.0734	0.0665
Q30	0.0071	0.0205	0.0011	0.0674
Q31	-0.0263	0.0205	0.0722	0.0663
Log likelihood function	-70,023.34		-5,694.127	
Restricted log likelihood	-82,072.15		-19,441.73	
Model test	$\chi^2 = 24{,}097.62$; df = 32; sig. = 0.0000		$\chi^2 = 27{,}495.20$; df = 33; sig. = 0.0000	
N	125,778		125,778	

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)
[2] Equation estimated on the entire sample
[3] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0)
* Significant at the 5% level
# Significant at the 10% level

Table A6.11.1. Results of participation and occupation selection equations estimated by independent probit

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	2.2252*	0.6064	0.7679*	0.0420
Years of education variables				
YED	0.0788*	0.0266	0.0534*	0.0052
YED2	-0.0024*	0.0009	-0.0007*	0.0002
Educational attainment variables				
NURSEQUA	-0.6494	0.4118	0.0976	0.0705
PGDEG	0.3069*	0.0843	0.2228*	0.0101
DEG	0.0572*	0.0261	0.1920*	0.0069
ALEVEL	-0.0640	0.0678	0.0817*	0.0063
NOQUAL	0.0741	0.1485	0.0350*	0.0071
Work experience variables				
EXP	0.0203*	0.0024	0.0267*	0.0007
EXP2	-0.0004*	0.0001	-0.0005*	0.00002
Personal characteristic variables				
DISABLE	0.1668*	0.0653	0.2867*	0.0120
ETHNIC	0.0575	0.0383	0.1476*	0.0101
NONBRIT	0.0755*	0.0363	0.1547*	0.0096
ETHNBRIT	0.0267	0.0650	-0.0932*	0.0203
Regional variables				
SEAST	0.0716*	0.0146	0.1561*	0.0034
Job characteristics variables				
HOURSPW	-0.0038*	0.0006	0.0019*	0.0001
MANAGE	0.0506*	0.0147	0.1386*	0.0038
NWORKERS	-0.0369*	0.0164	0.1338*	0.0033
TEMP	-0.0764*	0.0272	0.0099#	0.0060
Time trend variables				
Q18	-0.0531	0.0349	-0.0087	0.0092

Q19	-0.0550	0.0365	-0.0069	0.0092
Q20	-0.0072	0.0370	0.0164#	0.0093
Q21	-0.0154	0.0396	0.0035	0.0093
Q22	-0.0077	0.0353	0.0073	0.0092
Q23	-0.0419	0.0382	0.0185*	0.0092
Q24	0.0299	0.0382	0.0424*	0.0092
Q25	0.0281	0.0374	0.0580*	0.0093
Q26	0.0133	0.0391	0.0607*	0.0093
Q27	0.0465	0.0357	0.0742*	0.0093
Q28	0.0833*	0.0326	0.1117*	0.0081
Q29	0.0554#	0.0325	0.1138*	0.0082
Q30	0.0792*	0.0321	0.1045*	0.0082
Q31	0.0924*	0.0323	0.1348*	0.0083
Selection bias variables				
λ(p)	-0.3620*	0.0990	-0.5734*	0.0162
λ(nu)	-0.3370*	0.1601	0.0469	0.0632
Adjusted R^2	0.1340		0.3123	
Model test	F(34, 3,426) = 16.74; p = 0.0000		F(34, 77,198) = 1,032.70; p = 0.0000	
N	3,461		77,233	

[1] Dependent variable is LNWAGE

* Significant at the 5% level

Significant at the 10% level

Table A6.11.2. Results of selection-bias corrected wage equation estimated in the independent probit model

Appendix 6.12. Identification of the wage equation in the bivariate probit model

Below we distinguish the identifying variables of the wage equations for the bivariate probit model. The participation or occupation selection equation is estimated as described in the text. The wage equation however is re-estimated in each instance with no exclusion restrictions. One or both of the property income variables (PENSION or NONLABY) is in each instance found to be statistically insignificant at conventional levels.

	Nurses		All Other Workers	
	β_n [1]	Std.Err.	β_o [1]	Std.Err.
Constant	0.2058	0.2471	-1.1906*	0.0498
YED	0.1250*	0.0272	0.1388*	0.0053
YED2	-0.0038*	0.0009	-0.0035*	0.0002
NURSEQUA	0.0193	0.1805	0.2216*	0.0664
PGDEG	0.2429*	0.0721	0.3292*	0.0091
DEG	0.1034*	0.0283	0.2662*	0.0063
ALEVEL	-0.0687	0.0676	0.0589*	0.0061
NOQUAL	-0.0790	0.1529	-0.1893*	0.0066
EXP	0.0193*	0.0024	0.0228*	0.0007
EXP2	-0.0004*	0.0001	-0.0003*	0.00002
DISABLE	-0.1004	0.0885	-0.1515*	0.0103
ETHNIC	-0.0335	0.0480	-0.0168#	0.0095
NONBRIT	0.0418	0.0427	0.0287*	0.0092
ETHNBRIT	0.0054	0.0642	-0.1468*	0.0201
SEAST	0.0710*	0.0145	0.1585*	0.0034
HOURSPW	-0.0037*	0.0006	0.0024*	0.0001
MANAGE	0.0504*	0.0147	0.1300*	0.0037
NWORKERS	-0.0329*	0.0165	0.1293*	0.0032
TEMP	-0.0752*	0.0272	0.0099#	0.0060
Q18	-0.0513	0.0348	-0.0072	0.0091
Q19	-0.0384	0.0361	0.0070	0.0091
Q20	-0.0217	0.0363	0.0158#	0.0091
Q21	-0.0302	0.0379	0.0191*	0.0091
Q22	0.0050	0.0354	0.0216*	0.0091
Q23	-0.0523	0.0371	0.0335*	0.0091
Q24	0.0235	0.0372	0.0565*	0.0091
Q25	0.0202	0.0370	0.0650*	0.0092
Q26	0.0007	0.0379	0.0748*	0.0092
Q27	0.0613#	0.0355	0.0875*	0.0092
Q28	0.0726	0.0322	0.0981*	0.0081
Q29	0.0619*	0.0322	0.1017*	0.0081
Q30	0.0856*	0.0320	0.1051*	0.0081
Q31	0.0999*	0.0321	0.1248*	0.0082
λ(p)	0.0907	0.1449	0.0514*	0.0063
λ(nu)	-0.0702	0.0759	-0.0643	0.0548
AGE	0.0339*	0.0071	0.0528*	0.0011

AGE2	-0.0004*	0.0001	-0.0006*	0.00001
PCHILD	0.0230	0.0233	0.0857*	0.0054
COHABIT	0.0352	0.0271	0.0699*	0.0054
MARRIED	-0.0358*	0.0173	-0.0136*	0.0043
PENSION	0.0020	0.0831	0.0394	0.0161
NONLABY	0.000008	0.000007	-0.000005*	0.000002
Adjusted R^2	0.1394		0.3204	
Model test	F(41, 3,419) = 14.66; p = 0.0000		F(41, 77,191) = 925.16; p = 0.0000	
N	3,461		77,233	

[1] Dependent variable is LNWAGE

* Significant at the 5% level

Significant at the 10% level

Table A6.12.1. Results of the wage equation for the bivariate probit with no exclusion restrictions

Appendix 6.13. Results of the participation and occupation selection equations in the bivariate probit selection model with censoring

As noted in the text ρ_v is not statistically significant at conventional levels.

	Participation equation		Occupation selection equation	
	δ [1]	Std.Err.	γ [2]	Std.Err.
Constant	-2.2913*	0.0749	-2.1053*	1.0213
NURSEQUA			2.9921*	0.1412
Age variables				
AGE	0.0739*	0.0025	0.0023	0.0191
AGE2	-0.0008*	0.00003	-0.0001	0.0002
Personal characteristic variables				
DISABLE	-1.0733*	0.0110	-0.0666	0.2520
ETHNIC	-0.4531*	0.0199	0.2808*	0.1075
NONBRIT	-0.3573*	0.0200	0.3769*	0.0800
ETHNBRIT	-0.0062	0.0391	-0.0367	0.1289
Family variables				
PCHILD	-0.0798*	0.0136	0.1406*	0.0458
COHABIT	0.2450*	0.0146	-0.0761	0.0691
MARRIED	0.0540*	0.0108	-0.1160*	0.0362
Property income variables				
PENSION	-0.6746*	0.0297	0.0648	0.1875
NONLABY	-0.00003*	0.000002	-0.00003	0.00002
Years of education variables				
YED	0.2013*	0.0073	-0.0211	0.0642
YED2	-0.0069*	0.0002	0.0004	0.0021
Educational attainment variables				
PGDEG	0.5524*	0.0271	-0.7580*	0.1027
DEG	0.3342*	0.0171	-0.0841	0.0744
ALEVEL	-0.0912*	0.0155	0.0675	0.0709
NOQUAL	-0.6184*	0.0108	-0.2639	0.2145
Regional variables				
SEAST	0.0155#	0.0086	-0.0561#	0.0302
Time trend variables				
Q18	0.0020	0.0231	0.0032	0.0856
Q19	0.0484*	0.0233	0.0435	0.0848
Q20	0.0038	0.0233	-0.0538	0.0889
Q21	0.0574*	0.0235	-0.1805*	0.0914
Q22	0.0565*	0.0233	-0.0086	0.0887
Q23	0.0553*	0.0234	-0.1185	0.0859
Q24	0.0524*	0.0233	-0.1128	0.0893
Q25	0.0364	0.0235	-0.0789	0.0880
Q26	0.0596*	0.0235	-0.1560#	0.0884
Q27	0.0576*	0.0235	0.0341	0.0865
Q28	-0.0179	0.0203	-0.0447	0.0763
Q29	-0.0350#	0.0204	0.0913	0.0749
Q30	0.0072	0.0205	0.0224	0.0757
Q31	-0.0260	0.0205	0.0843	0.0744

ρ_v	-0.2354	
Log likelihood function	-75,143.82	
N	125,778	80,694

[1] Dependent variable is whether the individual participates in the labour market (PART = 1) or not (PART = 0)

[2] Dependent variable is whether the individual is a nurse (NURSE = 1) or not (NURSE = 0). This only observed when PART = 1.

* Significant at the 5% level

Significant at the 10% level

Table A6.13.1. Results of participation and occupation selection equations estimated jointly by bivariate probit with censoring

Appendix 6.14. Additional results of the decomposition analysis

We present in Table A6.14.1 further results of the decomposition analysis using: OLS estimates without correction for selection bias (based on the results of the statistical model presented in Appendix 6.6); participation selection bias corrected estimates (Appendix 6.7); occupation selection bias corrected estimates (Appendix 6.8 and 6.8); and, double selectivity corrected estimates estimated using an independent probit model (Appendix 6.11).

	Premium to being a nurse analysed using characteristics of nurses	Premium to being a nurse analysed using characteristics of non-nurses
OLS estimates		
Differences in variables (= differences in endowments)	0.3017	0.2732
Premium (= differences in returns to endowments)	0.0572	0.0857
Observed difference in mean lnW	0.3589	0.3589
Participation selection bias corrected estimates		
Differences in variables (= differences in endowments)	0.2609	0.2732
Premium (= differences in returns to endowments)	-0.1724	-0.1848
Differences due to participation selection bias	0.2704	0.2704
Observed difference in mean lnW	0.3589	0.3589
Occupation selection bias corrected estimates (version 1) [1]		
Differences in variables (= differences in endowments)	0.2934	0.1429
Premium (= differences in returns to endowments)	0.1132	0.2637
Differences due to occupation selection bias	-0.0477	-0.0477
Observed difference in mean lnW	0.3589	0.3589
Occupation selection bias corrected estimates (version 2) [2]		
Differences in variables (= differences in endowments)	0.3610	0.2087
Premium (= differences in returns to endowments)	0.0154	0.1678
Differences due to occupation selection bias	-0.0175	-0.0175
Observed difference in mean lnW	0.3589	0.3589
Independent probit model		
Differences in variables (= differences in endowments)	0.2139	-0.5492
Premium (= differences in returns to endowments)	0.2614	1.0245
Differences due to occupation and participation selection bias	-0.1164	-0.1164
Observed difference in mean lnW	0.3589	0.3589

[1] Occupation selection equation estimated using data for workers only (see Appendix 6.8)
[2] Occupation selection equation estimated using data for workers and non-workers (see Appendix 6.9)

Table A6.14.1. Additional results of decomposition analysis

Appendix 6.15. Detailed results of the decomposition analysis

	Premium to being a nurse analysed using characteristics of nurses		Premium to being a nurse analysed using characteristics of non-nurses	
	Premium	Differences in variables	Premium	Differences in variables
CONSTANT	0.6314	-	0.6314	-
YED	0.0748	0.0258	0.0733	0.0272
YED2	-0.1611	-0.0120	-0.1570	-0.0161
NURSEQUA	0.1282	0.1076	0.0034	0.2323
PGDEG	-0.0031	-0.0087	-0.0074	-0.0044
DEG	-0.0172	-0.0036	-0.0196	-0.0011
ALEVEL	-0.0009	-0.0030	-0.0077	0.0037
NOQUAL	0.0001	0.0144	0.0037	0.0108
EXP	-0.1036	0.0766	-0.0786	0.0517
EXP2	0.0198	-0.0346	0.0125	-0.0273
DISABLE	0.0021	0.0001	0.0024	-0.0002
ETHNIC	0.0003	0.0003	0.0002	0.0004
NONBRIT	0.0010	0.0015	0.0006	0.0019
ETHNBRIT	0.0018	-0.0008	0.0010	0.0000
SEAST	-0.0259	-0.0058	-0.0293	-0.0023
HOURSPW	-0.1876	0.0040	-0.1755	-0.0081
MANAGE	-0.0676	0.0712	-0.0225	0.0261
NWORKERS	-0.1428	0.0270	-0.1082	-0.0076
TEMP	-0.0047	-0.0001	-0.0062	0.0014
Q18	-0.0031	-0.0001	-0.0027	-0.0004
Q19	-0.0026	0.0000	-0.0027	0.0000
Q20	-0.0023	0.0000	-0.0023	0.0000
Q21	-0.0031	-0.0001	-0.0034	0.0002
Q22	-0.0016	0.0001	-0.0015	0.0000
Q23	-0.0052	-0.0001	-0.0054	0.0001
Q24	-0.0025	-0.0002	-0.0027	0.0000
Q25	-0.0028	-0.0002	-0.0029	0.0000
Q26	-0.0047	-0.0002	-0.0050	0.0000
Q27	-0.0023	0.0006	-0.0020	0.0004
Q28	-0.0041	-0.0011	-0.0045	-0.0007
Q29	-0.0048	0.0002	-0.0047	0.0001
Q30	-0.0036	-0.0002	-0.0037	-0.0001
Q31	-0.0036	0.0009	-0.0034	0.0007
Sub-total	0.0986	0.2597	0.0696	0.2887
Differences due to selection bias	0.0006		0.0006	
Observed difference in mean lnW	0.3589		0.3589	

Table A6.15.1. Bivariate probit selection model

	Premium to being a nurse analysed using characteristics of nurses		Premium to being a nurse analysed using characteristics of non-nurses	
	Premium	Differences in variables	Premium	Differences in variables
CONSTANT	0.7319	-	0.7319	-
YED	0.4841	0.0137	0.4750	0.0229
YED2	-0.3830	-0.0035	-0.3733	-0.0132
NURSEQUA	-0.1393	0.0400	-0.0037	-0.0956
PGDEG	0.0000	-0.0053	0.0000	-0.0053
DEG	-0.0110	-0.0024	-0.0126	-0.0009
ALEVEL	-0.0013	-0.0052	-0.0106	0.0041
NOQUAL	-0.0004	-0.0046	-0.0151	0.0101
EXP	-0.0653	0.0665	-0.0496	0.0508
EXP2	0.0064	-0.0286	0.0041	-0.0263
DISABLE	-0.0089	-0.0023	-0.0101	-0.0011
ETHNIC	-0.0050	0.0019	-0.0038	0.0006
NONBRIT	-0.0042	0.0034	-0.0027	0.0019
ETHNBRIT	0.0017	-0.0007	0.0010	0.0001
SEAST	-0.0228	-0.0056	-0.0259	-0.0025
HOURSPW	-0.1929	0.0042	-0.1804	-0.0083
MANAGE	-0.0647	0.0687	-0.0216	0.0256
NWORKERS	-0.1423	0.0269	-0.1079	-0.0076
TEMP	-0.0049	-0.0002	-0.0064	0.0014
Q18	-0.0030	-0.0001	-0.0027	-0.0004
Q19	-0.0024	0.0000	-0.0025	0.0000
Q20	-0.0019	0.0000	-0.0019	0.0000
Q21	-0.0018	0.0000	-0.0019	0.0002
Q22	-0.0009	0.0000	-0.0008	0.0000
Q23	-0.0041	0.0000	-0.0042	0.0001
Q24	-0.0014	-0.0002	-0.0015	-0.0001
Q25	-0.0019	-0.0001	-0.0020	0.0000
Q26	-0.0033	-0.0002	-0.0034	0.0000
Q27	-0.0016	0.0005	-0.0014	0.0003
Q28	-0.0032	-0.0010	-0.0035	-0.0007
Q29	-0.0052	0.0002	-0.0051	0.0001
Q30	-0.0025	-0.0002	-0.0026	-0.0001
Q31	-0.0037	0.0008	-0.0034	0.0006
Sub-total	0.1413	0.1668	0.3513	-0.0432
Differences due to selection bias	0.0508		0.0508	
Observed difference in mean lnW	0.3589		0.3589	

Table A6.15.2. Independent probit selection model with censoring

	Premium to being a nurse analysed using characteristics of nurses		Premium to being a nurse analysed using characteristics of non-nurses	
	Premium	Differences in variables	Premium	Differences in variables
CONSTANT	0.1616	-	0.1616	-
YED	0.4623	0.0176	0.4536	0.0264
YED2	-0.3589	-0.0064	-0.3498	-0.0155
NURSEQUA	0.0501	0.1559	0.0013	0.2047
PGDEG	-0.0020	-0.0072	-0.0048	-0.0044
DEG	-0.0142	-0.0031	-0.0162	-0.0010
ALEVEL	-0.0011	-0.0039	-0.0089	0.0040
NOQUAL	-0.0004	0.0066	-0.0152	0.0214
EXP	-0.0865	0.0721	-0.0657	0.0513
EXP2	0.0132	-0.0320	0.0083	-0.0272
DISABLE	-0.0052	-0.0012	-0.0059	-0.0006
ETHNIC	-0.0022	0.0011	-0.0017	0.0006
NONBRIT	-0.0012	0.0025	-0.0008	0.0021
ETHNBRIT	0.0015	-0.0006	0.0008	0.0001
SEAST	-0.0248	-0.0057	-0.0282	-0.0023
HOURSPW	-0.1897	0.0041	-0.1774	-0.0082
MANAGE	-0.0666	0.0703	-0.0222	0.0259
NWORKERS	-0.1415	0.0269	-0.1072	-0.0073
TEMP	-0.0048	-0.0002	-0.0064	0.0014
Q18	-0.0030	-0.0001	-0.0027	-0.0004
Q19	-0.0024	0.0000	-0.0024	0.0000
Q20	-0.0022	0.0000	-0.0022	0.0000
Q21	-0.0027	-0.0001	-0.0030	0.0002
Q22	-0.0013	0.0001	-0.0012	0.0000
Q23	-0.0048	-0.0001	-0.0050	0.0001
Q24	-0.0022	-0.0002	-0.0023	0.0000
Q25	-0.0025	-0.0001	-0.0026	0.0000
Q26	-0.0042	-0.0002	-0.0045	0.0000
Q27	-0.0019	0.0006	-0.0017	0.0004
Q28	-0.0038	-0.0010	-0.0042	-0.0007
Q29	-0.0048	0.0002	-0.0047	0.0001
Q30	-0.0031	-0.0002	-0.0031	-0.0001
Q31	-0.0036	0.0009	-0.0033	0.0007
Sub-total	-0.2529	0.2968	-0.2276	0.2715
Differences due to selection bias	0.3150		0.3150	
Observed difference in mean lnW	0.3589		0.3589	

Table A6.15.3. Trinomial logit selection model

	Premium to being a nurse analysed using characteristics of nurses		Premium to being a nurse analysed using characteristics of non-nurses	
	Premium	Differences in variables	Premium	Differences in variables
CONSTANT	0.4700	-	0.4700	-
YED	-0.2028	0.0330	-0.1989	0.0291
YED2	-0.0430	-0.0165	-0.0419	-0.0176
NURSEQUA	0.0777	0.1203	0.0021	0.1960
PGDEG	-0.0032	-0.0099	-0.0075	-0.0056
DEG	-0.0163	-0.0038	-0.0187	-0.0014
ALEVEL	-0.0088	-0.0728	-0.0719	-0.0097
NOQUAL	-0.0034	-0.1248	-0.1284	0.0001
EXP	-0.0577	0.0656	-0.0438	0.0517
EXP2	0.0063	-0.0300	0.0040	-0.0276
DISABLE	0.0025	0.0008	0.0029	0.0004
ETHNIC	-0.0009	0.0003	-0.0007	0.0001
NONBRIT	-0.0003	0.0015	-0.0002	0.0014
ETHNBRIT	0.0022	-0.0010	0.0012	0.0000
SEAST	-0.0238	-0.0056	-0.0270	-0.0024
HOURSPW	-0.2016	0.0051	-0.1885	-0.0080
MANAGE	-0.0652	0.0695	-0.0217	0.0260
NWORKERS	-0.1436	0.0275	-0.1088	-0.0073
TEMP	-0.0044	0.0000	-0.0058	0.0015
Q18	-0.0036	0.0000	-0.0032	-0.0004
Q19	-0.0029	0.0000	-0.0029	0.0000
Q20	-0.0032	0.0000	-0.0032	0.0000
Q21	-0.0028	-0.0001	-0.0031	0.0002
Q22	-0.0013	0.0001	-0.0012	0.0000
Q23	-0.0054	-0.0001	-0.0056	0.0001
Q24	-0.0029	-0.0003	-0.0031	-0.0001
Q25	-0.0024	-0.0002	-0.0025	-0.0001
Q26	-0.0047	-0.0003	-0.0049	0.0000
Q27	-0.0021	0.0007	-0.0019	0.0005
Q28	-0.0037	-0.0011	-0.0040	-0.0007
Q29	-0.0042	0.0002	-0.0042	0.0001
Q30	-0.0034	-0.0002	-0.0035	-0.0001
Q31	-0.0028	0.0009	-0.0027	0.0007
Sub-total	-0.2619	0.0589	-0.4299	0.2270
Differences due to selection bias	0.5619		0.5618	
Observed difference in mean lnW	0.3589		0.3589	

Table A6.15.4. Quadrinomial logit selection model

REFERENCES

Abdulai A. and Delgado CL. Determinants of non-farm earnings of farm-based husbands and wives in Northern Ghana. American Journal of Agricultural Economics 1999; 81: 117-30.

Adamache KW. and Sloan FA. Unions and hospitals: some unresolved issues. Journal of Health Economics 1982; 1: 81-108.

Ahlburg DA. and Mahoney CB. The effect of wages on the retention of nurses. Canadian Journal of Economics 1996; 29: S126-S129.

Amemiya T. Advanced econometrics. Basil Blackwell: Oxford, 1985.

Ault DE. and Rutman GL. Factors affecting levels and growth rates in the wage rates of women: evidence from nursing. Applied Economics 1998; 30: 727-39.

Baimbridge M. and Simpson C. Rewards to academia: the remuneration of vice chancellors and principals. Applied Economics 1996; 28: 631-639.

Becker GS. Human Capital. NBER: New York, 1964.

Beishon S., Virdee S. and Hagell A. Nursing in a multi-ethnic NHS. Policy Studies Institute: London, 1995.

Benham L. The labor market for registered nurses: a three equation model. Review of Economics and Statistics 1971; 53: 246-252.

Birch DM. and Calvert JR. How profitable is teaching? Higher Education Review 1973; 9: 35-44.

Bishop C. Manpower policy and the supply of nurses. Industrial Relations 1973; 12: 86-94.

Boganno MF., Hixson JS. and Jeffers JR. The short-run supply of nurses' time. Journal of Human Resources 1974; 9: 80-94.

Booton LA. and Lane JI. Hospital market structure and the return to nursing education. Journal of Human Resources 1985; 20: 184-96.

Bosworth D., Dawkins P. and Stronback T. The economics of the labour market. Longman; Harlow, 1996.

Botelho A., Jones CB. and Kiker BF. Nursing wages and educational credentials: the role of work experience and selectivity bias. Economics of Education Review 1998; 17: 297-306.

Bruggink T., Finan K., Gendel E. and Todd J. Direct and indirect effects of unionisation on the wage level of nurses: a case study of New Jersey hospitals. Journal of Labor Research 1985; 6: 405-416.

Buchan J. Flexibility of fragmentation: trends and prospects in nurses' pay. King's Fund Institute Briefing Paper No. 13. King's Fund Institute: London, 1995.

Buchan J. and Seccombe I. Nurse turnover costs: a review for the Royal College of Nursing. Institute for Employment Studies: Brighton, 1991.

Burstein PL. and Cromwell J. Relative rates of return for US physicians. Journal of Health Economics 1985; 4: 63-78.

Cabezas V. and MacDonald JB. Hysteresis and the earnings of immigrants in the United States labour market. Applied Economics 1999; 31: 1171-1182.

Carr-Hill RA., Dixon P., Griffiths M. et al. The impact of nursing grade on the quality and outcome of nursing care. Health Economics 1995; 4: 57-72.

Chow G. Tests of equality between sets of co-efficients in two linear regressions. Econometrica 1960; 28: 591-605.

Committee of Inquiry into the Pay and Conditions of Service of Nurses and Midwives (Chairman: Lord Halsbury), Report of the Committee of Inquiry into the Pay and Conditions of Service of Nurses and Midwives. HMSO: London, 1974.

Committee of Vice Chancellors and Principals and Universities Funding Council. University Management Statistics and Performance Indicators in the UK. CVCP: London, selected years.

Curti M. Does economic stagnation affect unemployed workers, even when re-employed? International Journal of Manpower 1998; 19: 410-23.

Day KM. and Devlin RA. Can volunteer work help explain the male-female earnings gap? Applied Economics 1997; 29: 707-721.

Department of Education. Higher Education Statistics for England. HMSO; London, selected years.

Department of Health and Social Services. Establishing priorities for health and personal social services in England. Department of Health and Social Services, London: 1976a.

Department of Health and Social Services. Sharing resources for health in England: report of the Resource Allocation Working Party. HMSO, London: 1976b.

Department of Health. Community care statistics, 2000. Department of Health, London: 2001.

Department of Health. NHS Hospital and Community Health Services Non-Medical Workforce Census. Department of Health, London: selected years.

Dolton PJ. and Makepeace GH. Interpreting sample selection effects. Economics Letters 1987; 27: 373-379.

Dolton PJ. and Makepeace GH. Sample selection and male-female earnings differentials in the graduate labour market. Oxford Economic Papers 1986; 38: 317-341.

Earle JS. and Sakova Z. Business start-ups or disguised employment? Evidence on the character of self-employment from transition economies. Labour Economics 2000; 7: 575-601.

Eckhaus RS. Estimation of the returns to education with hourly standardised incomes. Quarterly Journal of Economics 1973; 87: 121-131.

Elliott RF. Labour economics: a comparative text. McGraw-Hill: London, 1991.

Elliott RF. and Duffus K. What has been happening to pay in the public service sector of the British economy: developments over the period from 1970 to 1992, Treasury Occasional Paper No. 3, HM Treasury, London: 1995. (Also published in British Journal of Industrial Relations 1996; 34: 51-85.

English National Board. ENB Annual Report. ENB: London, selected years.

Feldman R. and Scheffler R. The union impact on hospital wages and fringe benefits. Industrial and Labor Relations Review 1982; 35: 196-206.

Freese J. and Long JS. Tests for the multinomial logit model. Stata Technical Bulletin 2000; 58: 19-25.

French MT. And Dunlap LJ. Compensating wage differentials for job stress. Applied Economics 1998; 30: 1067-1066.

Goodin RE. Discounting discounting. Journal of Public Policy 1976; 2: 53-72.

Government Actuary Department. English Life Tables no.14 Series DS no.7. HMSO: London, 1992.

Gray AM. The NHS and the history of nurses' pay. Bulletin of History of Nursing 1989; 2: 15-29.

Gray AM. and Normand C. The hidden costs of turnover. Health Service Journal 1990; 100: 1646-1647.

Gray AM. and Phillips VL. Local turnover in the British National Health Service: a local labour market analysis. Health Policy 1996; 36: 117-128.

Greene WH. Econometric analysis (fourth edition). Prentice Hall: New York, 2000.

Greene WH. LIMDEP, version 7.0 Users Manual. Econometric Software: New York, 1998.

Gronau R. Wage comparisons – a selectivity bias. Journal of Political Economy 1974; 82: 1119-43.

Groot W. and Mekkeholt E. The rate of return top investments in on-the-job training. Applied Economics 1995; 27: 173-181.

Groot W. and Van Den Brink HM. Education, training and employability. Applied Economics 2000; 32: 573-582.

Groot W. The incidence of, and returns to overeducation in the UK. Applied Economics 1996; 28: 1345-1350.

Gujarati DN. Basic econometrics (second edition). McGraw-Hill: London, 1988.

Gyimah-Brempong K. and Fichtenbaum R. Racial wage gaps and differences in human capital. Applied Economics. 1997; 29: 1033-1044.

Gyourko J. and Tracey JS. An analysis of public and private sector wages allowing for endogenous choices of both government and union status. National Bureau of Economic Research Working Paper 1920. NBER: New York, 1986.

Halvorsen R. and Palmqvist R. The interpretation of dummy variables in semilogarithmic equations. American Economic Review 1980; 70: 474-475.

Hamilton BH. Racial discrimination and professional basketball salaries in the 1990s. Applied Economics 1997; 29: 287-296.

Hancock C. Report of the RCN Autumn conference, October 20th 1999. Secretary General of the RCN.

Harris RI. What is the hourly wage differential between full-time and part-time workers? Applied Economics 1993; 25: 1-12.

Hausman J. and McFadden D. Specification tests in econometrics. Econometrica 1984; 52: 1219-40.

Heckman JJ. Sample selection bias as a specification area. Econometrica 1979; 47: 153-161.

Hirsch BT. and Schumacher EJ. Monopsony power and relative wages in the labor market for nurses. Journal of Health Economics 1995; 14: 443-76.

Hurd R. Equilibrium vacancies in a labour market dominated by non-profit firms: the shortage of nurses. Review of Economics and Statistics 1973; 55: 234-240.

James E., Alsalam N., Conaty JC. and To DL. College quality and future earnings: where should you send youor child to college? American Economic Review (Papers and Proceedings) 1989; 79: 247-252.

Jenkins-Clarke S. Measuring nursing workload: a cautionary tale. Centre for Health Economics Discussion Paper 96. Centre for Health Economics: University of York, 1992.

Johnes G. Schooling, fertility and the labour market experience of married women. Applied Economics 1999; 31: 585-592.

Jones JC., Nadeau S. and Walsh WD. Ethnicity, productivity and salary: player compensation and discrimination in the National Hockey League. Applied Economics 1999; 31: 593-608.

Joshi H. and Paci P. Unequal pay for women and men: evidence from the British Birth Cohort Studies. The MIT Press: Cambridge, Massachusetts, 1998.

Kanellopoulos CN. Public-private wage differentials in Greece. Applied Economics 1997; 29: 1023-1032.

Killingsworth MR. Labour supply. CUP: Cambridge, 1983.

Lambert SF. Labour market experience in female wage equations: does the experience measure matter? Applied Economics 1993; 25: 1439-1449.

Lanzona LA. Migration decisions, self-selection and earnings in low-income Philippine communities. Yale Economic Growth Center Discussion Paper 747. Yale University: New York, 1995.

Lee LF. Generalized econometric models with selectivity. Econometrica 1983; 51: 507-512.

Lehrer EL. White WD. and Young WB. The three avenues to registered nurse license: a comparative analysis. Journal of Human Resources 1991; 26: 362-379.

Lewis HG. Comments on selectivity bias in wage comparisons. Journal of Political Economy 1974; 82: 1145-1155.

Lindsay CM. More real returns to medical education. Journal of Human Resources 1976; 11: 127-130.

Lindsay CM. Real returns to medical education. Journal of Human Resources 1973; 8: 331-348.

Link CR. and Landon J. Monopsony and union power in the market for nurses. Southern Economic Journal 1975; 41: 649-659.

Link CR. and Settle RF. A simultaneous-equation model of labor supply, fertility and earnings of married women: the case of registered nurses. Southern Economic Journal 1979; 47: 977-89.

Link CR. and Settle RF. Wage incentives and married professional nurses: a case of backward-bending supply? Economic Inquiry 1981; 19: 144-56.

Link CR. Returns to nursing education: 1970-1984. Journal of Human Resources 1988; 23: 372-387.

Maglen L. and Layard R. How profitable is engineering? Higher Education Review 1970; 2: 51-67.

Manski C. Non-parametric bounds on treatment effects. American Economic Review 1990; 80: 319-323.

McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P. (ed.) Frontiers in econometrics. Academic Press: New York, 1973.

McLennan MC. Does household labour impact market wages? Applied Economics 2000. 32: 1541-1557.

Mennemeyer ST. and Gaumer G. Nursing wages and the value of educational credentials. Journal of Human Resources 1983; 18: 32-48.

Metcalf, D. The rate of return to investing in a doctorate: a case study. Scottish Journal of Political Economy 1973; 20: 43-51.

Mincer J. Schooling, experience and earnings. NBER: New York, 1974.

Moffitt RA. New developments in econometric methods for labor market analysis. In: Ashenfelter O. and Layard R. (ed.s) Handbook of Labour Economics, Volume 3. Elsevier: New York, 1999.

Mohanty MS. Determination of participation decision, hiring decision, and wages in a double selection framework: male-female wage differentials in the U.S. labor market revisited. Contemporary Economic Policy 2001; 19: 197-212.

Moretti E. Do wages compensate for risk of unemployment? Parametric and semiparametric evidence from seasonal jobs. Journal of Risk and Uncertainty 2000; 20: 45-66.

Morris S. Health economics for nurses: an introductory guide. Prentice Hall: London, 1998.

Morris V. and Ziderman A. The economic return on investment in higher education in England and Wales. Economic Trends 1971; 211: 20-31.

Mott DA. and Kreling DH. An internal rate of return approach to investigate pharmacist supply in the United States. Health Economics 1994; 3: 373-84.

Netten A., Knight J., Dennett J., Cooley R. and Slight A. A ready reckoner for staff costs in the NHS. PSSRU: Canterbury, 1998.

Oaxaca R. Male-female wage differentials in urban labour markets. International Economic Review 1973; 14: 693-709.

Office of Health Economics. Compendium of health statistics, 2001. OHE: London, 2001.

Ometto AH., Hoffman R. and Correa-Alves M. Particppacao da mulher no mercado de trabalho: discriminacao em Pernambuco e Sao Paulo (with English summary). Revista Brasileira de Economia 1999; 53: 287-322.

ONS. New Earnings Survey, Part D: analyses by occupation. The Stationery Office: London, selected years.

ONS. Annual Abstract of Statistics. The Stationery Office: London, selected years.

ONS. Health and Personal Social Services Statistics for England. The Stationery Office: London, selected years.

ONS. UK National Accounts. The Stationery Office: London, selected years.

OPCS (1995) Qualified nurses, midwives and health visitors, HMSO, London.

Pagan JA. and Davila A. Obesity, occupational attainment, and earnings. Social Science Quarterly 1997; 78: 756-770.

Pagan JA. and Tijerina-Guajardo JA. Increasing wage dispersion and the changes in relative employment and wages in Mexico's urban informal sector: 1987-1993. Applied Economics 2000; 32: 335-347.

Palme MO. and Wright RE. Changes in the rate of return to education in Sweden: 1968-1991. Applied Economics 1998; 30: 1653-1664.

Parker SC. Hystersis and the earnings function. Applied Economics 1994; 26: 465-468.

Pearce DW. and Nash CA. The social appraisal of projects. MacMillan: Hong Kong, 1981.

Phillips VL. Nurses' labor supply: participation, hours of work, and discontinuities in the supply function. Journal of Health Economics 1995; 14: 567-82.

Pissarides, CA. From school to university: the demand for post-compulsory education in Britain. Economic Journal 1982; 92: 654-67.

Polachek SW. and Siebert WS. The economics of earnings. CUP: Cambridge, 1993.

Prescott D. and Wandschneider B. Public/private sector wage differentials in Canada – evidence from the 1991 and 1982 surveys of consumer finance. Applied Economics 1999; 31: 723-731.

Psacharopoulos G. Returns to education. Journal of Human Resources 1985; 14: 584-604.

Psacharopoulos G. and Layard PR Human capital and earnings: British evidence and a critique. Review of Economic Studies 1979; 46: 485-503.

Psacharopoulos G. Rates of return to investment in education around the world. Comparative Education Review 1972; 16: 54-67.

Psacharopoulos G. Returns to education: an international comparison. Elsevier: Amsterdam, 1973.

Psacharopoulos G. Returns to education: an updated comparison. Comparative Education 1981; 17: 321-41.

Puhani PA. The Heckman correction for sample selection and its critique. Journal of Economic Surveys 2000; 14: 53-68.

Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 1998. Fifteenth report on nursing staff, midwives and health visitors. HMSO: London, 1998.

Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 1984. First report on nursing staff, midwives and health visitors 1984. HMSO: London, 1984.

Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 2001. Eighteenth report on nursing staff, midwives and health visitors. The Stationery Office: London, 2001.

Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine, 1992. Ninth report on nursing staff, midwives and health visitors. HMSO: London, 1992.

Richardson G. and Maynard A. Fewer doctors? More nurses? A review of the knowledge base of doctor-nurse substitution. Centre for Health Economics Discussion Paper 135. Centre for Health Economics: University of York, 1995.

Robst J. and Cuson-Graham K. The effect of uncertain educational requirements on education and wages. Applied Economics 1999; 31: 53-63.

Sanmartin M. Linearity of the returns to education and self-selection. Applied Economics 2001; 33: 133-142.

Schuld TC., Schippers JJ. and Siegers JJ. Allocation and wage structure: differences between men and women. Applied Economics 1994; 26: 137-152.

Schumacher EJ. and Hirsch BT. Compensating differentials and unmeasured ability in the labor market for nurses: why do hospitals pay more? Industrial and Labor Relations Review 1997; 50: 557-79.

Seccombe I. And Smith G. Taking part: registered nurses and the labour market in 1997. Institute for Employment Studies: Brighton, 1997.

Shields MA. and Ward M. Improving nurse retention in the National Health Service in England: the impact of job satisfaction on intentions to quit. Journal of Health Economics 2001; 20: 677-701.

Shields MA. And Wheatley Price S. The earnings of male immigrants in England: evidence from the quarterly LFS. Applied Economics 1998; 30: 1157-1168.

Sicilian P. and Grossberg AJ. Investment in human capital and gender wage differences: evidence from NLSY. Applied Economics 2001; 33: 463-472.

Siebert WS. Occupational licensing: the Merrison Report on the regulation of the medical profession. British Journal of Industrial Relations 1977; 15: 39-41.

Simpson W. Intermittent work activity and earnings. Applied Economics 2000; 32: 1777-1786.

Sloan F. and Richupan S. Short-run supply responses of professional nurses: a microanalysis. Journal of Human Resources 1975; 10: 241-257.

Small KA. and Hsiao C. Multinomial logit specification tests. International Economic Review 1985; 26: 619-27.

Sorensen E. Continuous female workers: how different are they from other women? Eastern Economic Journal 1993; 19: 15-32.

Sorensen E. Measuring the pay disparity between typically female occupations and other jobs: a bivariate selectivity approach. Industrial and Labor Relations Review 1989; 42: 624-39.

Standing Commission on Pay Comparability (Chairman: Clegg, MH.), 1980. Report of the Standing Commission on Pay Comparability, Report No.3, Nurses and Midwives. HMSO: London, 1980.

Sugden R. and Williams A. The principles of practical cost-benefit analysis. Oxford University Press: London, 1978.

Sullivan D. Monopsony power in the market for nurses. Journal of Law and Economics 1989; 32: S135-78.

Tanaka Y. Employment tenure, job expectancy, and earnings profile in Japan. Applied Economics 2001; 33: 365-374.

Thompson MS. Benefit-cost analysis for program evaluation. Sage: London, 1980.

Tomes N. Religion and the earnings function. American Economic Review (Papers and Proceedings) 1985; 75: 245-250.

Tunali I. A general structure for models of double selection and an application to a joint migration/earnings process with remigration. In Ehrenberg RG. (ed.) Research in Labor Economics Volume 8 (Part B). JAI Press. Greenwich, Connecticut, and London, 1986.

Vella F. Estimating models with sample selection bias: a survey. Journal of Human Resources 1998; 33: 127-169.

White H. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica 1980; 48: 817-38.

Wilkinson BW. Present values of lifetime earnings for different occupations. Journal of Political Economy 1966; 74: 556-572.

Williams G. and Gordon A. Perceived earnings functions and ex ante rates of return to post compulsory education in England. Higher Education 1981; 10: 199-227.

Willis RJ. Wage determinants, a survey and reinterpretation of human capital earnings functions. In: Ashenfelter O. and Layard R. (eds.) Handbook of Labour Economics, Volume 1. Elsevier: New York, 1986.

Wilson RA. A longer perspective on rates of return. Scottish Journal of Political Economy 1985a; 32: 191-8.

Wilson RA. Rates of return to entering the legal profession: some further evidence. Scottish Journal of Political Economy 1987a; 34: 174-91.

Wilson RA. Rates of return: some further results. Scottish Journal of Political Economy 1983a; 30: 114-27.

Wilson RA. Returns to entering the medical profession in the UK. Journal of Health Economics 1987b; 6: 339-49.

Wilson RA. The declining returns to becoming a teacher. Higher Education Review 1983b; 15: 22-38.

Wilson RA. The rate of return to becoming a qualified scientist or engineer in Great Britain, 1966-1976. Scottish Journal of Political Economy 1980; 21: 41-62.

Wilson RA. The return on entering the business professions. Business Economist 1984; 16: 30-41.

Wilson RA. The social returns to producing teachers. Higher Education Review 1985b; 17: 59-69.

Yett DE. The nursing shortage. In: Cooper MH. And Culyer AJ (ed.s) Health economics: selected readings. Penguin: Middlesex, 1973.

Ziderman, A. Does it pay to take a degree? The profitability of private investment in university education in Britain. Oxford Economic Papers 1973; 15: 262-274.

Zweimuller J. Survey non-response and biases in wage regressions. Economics Letters 1992; 39: 105-9.