APPROACHES TO THE MEASUREMENT OF EFFICIENCY IN A DYNAMIC CONTEXT

WITH AN APPLICATION TO THE UK BANKING SECTOR

JONATHAN SANDBACH

CITY UNIVERSITY

MARCH 1993

A Thesis submitted for the degree of Ph.D.

COPYING

ACKNOWLEDGEMENTS

I would like to thank Dr. P. Pashardes and Prof. D. Glycopantis for their time and supervision during the course of my research.

ABSTRACT

In this thesis we put forward two approaches to the measurement of the cost efficiency of firms in which adjustment costs affect the optimal allocation of factor inputs. The two approaches we consider are, first, based on direct calculation of an efficiency index and, second, the estimation of a parametric model. The implementation of both approaches require extensions to the existing theory.

In developing the cost efficiency index approach we suggest that the Tornqvist index can be extended to be "exact and superlative" (Diewert (1976)), even when there are adjustment costs. However, the formulae for this index becomes dependent on the unknown adjustment cost function and this needs to be estimated in reduced form by econometric methods.

We find an extension of the parametric model approach that incorporates adjustment costs to be preferable, since it gives a greater understanding of the structure of adjustment costs - not only their direct influence on the costs of firms, but also the extent to which they alter the optimal factor demands (a feature that in a static model might be interpreted as allocative inefficiency).

The empirical analysis in this thesis applies both approaches to estimate efficiency differences between firms, and over time, in the UK retail banking and building society sectors. The parametric model we develop, incorporating adjustment costs, is particularly appropriate since these sectors have undergone considerable change over the period we study (1978 to 1987), in terms of both the level of output and optimal factor demands (with labour substituted for progressively more computer and information technology equipment). Furthermore, the parametric approach is easily adapted to deal with other particular characteristics of these sectors. In particular, we are able to extend our model to freely estimate equipment depreciation rates. This results in using an unbiased estimate of the user cost of equipment when estimating cost efficiency.

CONTENTS

LIST OF TABLES, CHARTS AND FIGURES

CHAPTER I

INTRODUCTION

ABSTRACT

This chapter discusses adjustment costs and their implications for the measurement of efficiency. It also introduces the two methods of cost efficiency measurement that will be developed in this thesis: an index number approach and a parametric model based approach. Finally, the chapter discusses the reasons why adjustment cost models may be particularly appropriate to the measurement of cost efficiency in the banking and building society sectors.

1. Adjustment Costs and Efficiency

This thesis is concerned principally with two approaches to comparing the cost efficiency of firms. These are, respectively, an index number approach and a parametric model based approach. Index number measurements of efficiency can be calculated directly from individual sets of observations on firms (or one firm at different times). They require no explicit knowledge of the underlying production or cost structure. However, under certain assumptions about the form of either the production or cost function, these index numbers can acquire an explicit "economic" interpretation. A parametric model approach requires explicit identification and estimation of a production or cost structure, against which the efficiency of any individual firm can be measured.

This thesis develops extensions of the index number and parametric approaches to situations where adjustment costs are included in a dynamic model of production. It argues that, while both approaches can incorporate adjustment costs, the parametric approach is more amenable to identification of an underlying dynamic structural model and, even, the firm's expectation forming process for future factor input requirements. This permits more insights into the causes of apparent cost and allocative inefficiencies.

Efficiency measurement is important to managers, investors and industrial policy makers alike. For example, managers wish to know how their production processes compare in productivity and unit costs to other similar companies or best practice amongst competitors. Similarly, investors wish to be able to assess the performance of their investments. Recently, public policy makers have become concerned with the efficiency improvements of regulated companies operating in dominant market positions. In these cases, competitive pressures cannot be relied upon to ensure a company's efficiency, and so a regulator must monitor the efficiency of the company in terms of its improvement over time and in relation to similar companies in other markets[1]. These are some of the reasons for which economists have long been interested in efficiency measurement, and reasons for apparent efficiency differences.

Firms incur adjustment costs when they change inputs in response to changes in outputs, factor prices or expectations of future output requirements and factor prices. In a static model, adjustments costs may not be separately identified and, therefore, will be treated as

---

1 For example, efficiency measurement was discussed in Oftel (1992). Oftel (Office of Telecommunications - the telecommunications industry regulator) has also decided that the efficiency of BT's access business should be assessed in relation to local telephone carriers in the USA (see BT's License amended in 1991). Similarly, Ofwat (Office of Water Supply - the regulator of the water supply industry) is developing a form of *benchmark* regulation for privatised water companies that relies heavily on comparative efficiency measurement.

inefficiencies. However, in a dynamic model, adjustment costs (or their impact on a firm's decisions) can be separately identified. In this case, we may be able to exclude from our measure of inefficiency adjustment costs that are unavoidable for the cost minimising firm (ie. any adjustment cost necessary for the firm to minimise the sum of adjustment costs and disequilibrium costs). However, part of the adjustment costs may also be considered as inefficiency, for example, when firms fail to lessen their exposure to adjustment costs by not concentrating changes on factor inputs that can be varied at a relatively lower cost. A further example would be a firm that failed to react in an optimal manner to rational expectations concerning future factor input requirements, thus failing to minimise the present value of current and future total costs (including adjustment costs). Again, any excess adjustment costs incurred as a result should be regarded as inefficiency.

This thesis contributes in two directions to the discussion of the effects of dynamic structure on efficiency measurement. First, it extends the Tornqvist cost efficiency index to deal with an underlying dynamic cost function, while retaining the *exact* and *superlative* properties as defined by Diewert (1976)[2]. Although the formulation of the index becomes dependent on an unknown adjustment cost function, we show that the reduced form dynamics of the index can be empirically estimated through econometric analysis. However, as we shall see, through this approach we cannot achieve a full understanding of the underlying structure and magnitude of adjustment costs.

In a second approach to the problem of efficiency measurement in the context of a dynamic cost structure, we explore the extension of a parametric method, based on behavioural relationships, to incorporate an adjustment cost function. We show how the behaviour of the firm is affected by such costs using a set of estimatable non-linear factor demand functions. These explicitly include parameters of the

---

2 *Exact* describes an index that can be analytically derived from an underlying function, and *superlative* describes an index that is *exact* for a *flexible* underlying economic function.

adjustment cost function. Estimates of these parameters are then used to "strip out" estimates of the dynamic effects from the observed costs and factor shares, leaving equilibrium values for costs and factor shares (ie. those that would apply in the absence of adjustment costs). These "actual" equilibrium values are then used in a second stage estimation of a system of equilibrium cost function and factor share equations, using output, factor prices, a time trend and quasi-fixed inputs as independent variables. The fitted values of the cost function in this system provides the benchmark from which cost inefficiencies are estimated.

A further advantage of the construction of a structural model is that it can be adapted to suit specific applications. We illustrate this by adapting the model to incorporate unknown depreciation rates for each input, which may differ from the depreciation rates assumed in the construction of the data.

2. Application to the Banking Sector

In this thesis we apply both approaches discussed above to efficiency measurement using data from the UK banking and building society sector between 1978 to 1987. A number of special factors characterise the bank and building sector of the UK over this period. Two of these factors are particularly relevant in the context of adjustment costs discussed in this thesis. They are, first, the exceptionally high rate of growth of the sectors (compared to the UK economy as a whole) and, second, the effect of technical change.

With respect to the growth of the sector, Central Statistical Office industry output statistics indicate that between 1978 and 1987 the banking sector grew by 87.7% in real terms (or an average of 7.2% per annum). This compares with real growth of 19.8% (or 2.0% per annum) for the whole economy. This rapid growth of the banking sector can be attributed to developments on both the "supply-side" and "demand-side".

On the supply side of the industry there have been new sources of competition for banks and building societies, and a loosening of a number of regulatory restrictions on lending. New competition has come from the entry of banks into mortgage lending markets and building societies into current account and other banking service markets, and the appearance of other new retail banking companies. Regulatory lending restrictions that have been abolished include guidelines on building society lending that restricted the supply of mortgages, Supplementary Special Deposits (the *corset*), Reserve Asset Ratio Requirements and hire purchase restrictions. These factors have all expanded the sectors's supply capability. Furthermore, service innovation (such as electronic banking) has expanded the sector's product portfolio.

On the demand side, growth in GDP and personal incomes has resulted in a more than proportionate rise in the demand for banking services. Like many service industries, both personal and corporate banking have national income elasticities greater than unity[3]. Despite the recession of the early 1980s, the UK economy grew markedly over the period as a whole, creating the environment for substantial growth in the banking sector.

The second characteristic of the banking sector that is relevant for us to consider is that of technical change. The banking sector is based on information handling (e.g. account and loan details) and is well suited to new information technologies (ie. computers and telecommunications). Central Statistical Office Input/Output Tables for 1984 indicate that over 70% of non-building investment by the banking sector was in computers and office machinery, and this

---

3 The strong dependency of banking on the health of the economy generally is shown by regressing the output of the sector against real consumers' expenditure and GDP (to capture the effects of economic growth on personal and corporate banking respectively). One finds that the sum of the GDP and consumers' expenditure elasticities is significantly greater than unity.

proportion has certainly risen sharply since the mid 1980s as major banks and building societies either embarked or continued with extensive computerisation programs. Examples include the following:

- Most visibly tellers in banks and building societies have often been replaced by ATMs (Automatic Teller Machines). Originally these only dispensed cash, but by the mid 1980s they also performed a range of other functions (for example, see National and Provincial (1986), pages 14 and 15). Hannan and McDowell (1987) have looked at a diffusion model to study the reasons and mechanisms by which ATMs have multiplied through the banking sector, but the effect of these changes on cost efficiency has not been properly quantified.

- The remaining tellers' positions have been installed with various input/output devices to link them to computer systems that perform operations previously performed manually by the tellers and back-office clerical staff (for example, see National and Provincial (1986), page 16).

- SWIFT and other messaging and transaction clearing systems replaced other back-office clerical positions.

- Centralised marketing databases opened up new possibilities for company expansion through the identification of cross-selling opportunities.

- Home banking and EFT (Electronic Funds Transfer) linking directly into the banks computer systems to a lesser degree also replaced tellers and back-office clerical work in branches.

- Building societies, which already averaged one computerised work station per two employees by the middle of the 1980s, accelerated their investment in information technology to grow by between 15% and 18% per annum by 1987. This investment was believed to have been concentrated on new mortgage systems to handle all the processing for the large variety of new products building societies now offer. (See, for example, National and Provincial (1986), pages 14 and 15, and International Data Corporation (1989)).

As the number of accounts grew, more transactions were generated and the scale of the information handling operation correspondingly increased, as did its consequent importance to banks and building societies (see, for example, the comments made in National and Provincial (1986), pages 15 and 16, regarding computer systems). It has even been claimed that there were some analysts who allotted ratings to banks on the basis of their investment in telecommunications and computers in addition to such standard criteria as profits, cash flow and reserves, presumably on the assumption that this investment is important to yielding good future returns (see British Telecommunications Engineering (1988)).

However, one recent study by Roach (1987) suggests that heavy investment did not yield any great productivity benefits. However, this conclusion is based on a simple comparison of productivity statistics with measures of the technological intensity of each sector. A survey by the International Data Corporation (1989) found that even the companies themselves had difficulty in measuring the contribution made by information technology, with decision makers describing investment decisions as being based on *gut feel* or being simply *strategic*, and very likely to be prompted by the fear that most other companies were investing heavily and that they risked being left behind. Therefore, on the basis of this evidence, the role of information technology in enabling productivity growth is far from clear.

Two principal implications, relevant to this thesis, can be drawn from this discussion of changing technology in banks and building societies. First, the rapidly changing technological environment meant that asset lives of plant and machinery were shortened, and so accounting estimates of fixed asset values tended to under-estimated their true user cost. Although a system could have physically lasted up to ten years, it may have become technologically obsolete in two to three years.

Second, the complexities of new information technology systems and the fundamental changes in processes that they implied, meant that we would expect banks and building societies to have faced considerable short-term adjustment costs during the first year of operation of new systems. Hardware and software bugs would have needed to be ironed out, staff trained and the bank or building society would have had to familiarise itself with the new system and its potential.

It can be seen, therefore, that in attempting to estimate the growth in efficiency of the bank and building society sectors, it is essential to consider the role adjustment costs play in the cost minimising combinations of labour and equipment that banks and building societies employ. It is also important to consider the effective depreciation rate for equipment used by banks (as opposed to its accounting counterparts).

Previous studies (e.g. Roach (1987)) have failed to find any productivity benefits from the considerable investment of this sector in computer and other information technology equipment. However, this thesis has found that adjustment costs are relatively lower for equipment (compared to labour) and this, therefore, should be one of the attractions of investment in computers and related equipment in a growing industry. Productivity in the banking and building society sectors estimated from a static model may appear low because of additional investment in information technology equipment. In fact it is this investment that has enabled the sector to cope with high output growth whilst minimising the adjustment costs that this growth entails.

3. Structure of the Thesis

The structure of this thesis is as follows. Chapter II reviews the existing literature on the subject of efficiency measurement (in both static and dynamic models). In order to compare adequately the different approaches adopted by the various authors and studies, we initially set out a common terminology and notation. We also consider which approaches in the literature can most suitably be adapted to take account of adjustment costs.

Chapters III and IV then describe in detail two dynamic models of the firm incorporating adjustment costs and formulate appropriate efficiency measures, using respectively an index number and parametric approach. Both approaches are illustrated by empirical analysis of the UK bank and building society sectors. The parametric model based approach allows for estimation of other behavioural characteristics. These include the expectation forming process for future factor input requirements and the average industry asset life of plant and machinery. This is preferable to relying on accounting assumptions which may not take adequate account of technological change making assets obsolete before the end of their physical lives (as is usually done in an index number approach).

Finally, chapter V presents the conclusions of this thesis by comparing the experience of applying the two approaches to efficiency measurement in the context of cost structures that incorporate adjustment costs. Furthermore, we draw conclusions concerning efficiency improvements and changes in the UK banking sector over the period under investigation, 1978 to 1987. We find that adjustment costs for the stock of both labour and equipment are important in determining the long run cost minimising factor demands in such a fast growing industry which has also seen a number of changes to its regulatory and competitive environment.

CHAPTER II

THE MEASUREMENT OF EFFICIENCY: INDICES, PROGRAMMING AND ECONOMETRICS

APPROACHES TO EFFICIENCY MEASUREMENT

ABSTRACT

We review the theory of efficiency measurement and methods for empirical application such as direct index measurement, linear or mathematical programming based methods (such as DEA) and the use of parametric models. We also review work to date on the measurement of efficiency in a dynamic context. We conclude by suggesting the most fruitful lines for further development of efficiency measurement in dynamic models to be direct index numbers and parametric models.

1. Introduction

In this chapter we critically review parametric and non-parametric approaches to the measurement of efficiency, describing their inter-relations and relative strengths and weaknesses.

Parametric approaches use observations from a sample of firms to estimate a behavioural model of the production or cost structure. The efficiency of each firm can then be measured relative to the estimated function (defining the reference technology). Some non-parametric methods also define a production or cost model by a convex hull around the observed data for the sample of firms (e.g. linear programming based methods). Although no parameters are estimated there is, nevertheless, an estimated benchmark (formed by a series of intersecting hyperplanes) from which efficiency can be measured. Other non-parametric methods do not even estimate any underlying behavioural model. Instead they use an index to directly calculate the difference in efficiency between two firms (or one firm and a reference, such as the *sample average*). An index can be given an economic interpretation provided that underlying production or cost functions of both firms are drawn from a general class of function in which unknown parameters *cancel out* in the formulation of the efficiency index (as we will describe in more detail later). We will refer to these latter methods as *direct efficiency indices.*

Although (as we shall see later) parametric and non-parametric approaches require assumptions concerning the underlying production and cost structures of the firms under analysis, only the parametric approach requires an explicit identification and estimation of this structure. This is usually done through an econometric model of the production or cost function. Non-parametric approaches are preferred by some investigators since they avoid the need to assume an explicit functional form for the production or cost function. On the other hand, directly calculated indices can often be difficult to interpret, whereas an estimated behavioural model can lead to a meaningful economic interpretation of the results with regard to efficiency.

A number of papers have been published that compare parametric techniques (regression based models) with non-parametric techniques (for example, linear programming and Data Envelope Analysis (DEA)), as in Lovell and Schmidt (1988). In this chapter we take a broader view of techniques available to the investigator. These include directly calculated efficiency indices (e.g. Total Factor Productivity (TFP) indices and, more generally, Tornqvist indices) in which no behavioural model is explicitly estimated. We also discuss behavioural models that incorporate dynamic characteristics into the firms' production and cost functions. We note that, as yet, the dynamic structure in a firm's production or cost function has not been incorporated into the underlying theory of directly calculated efficiency indices. This is the subject of a later chapter.

This chapter is organised as follows. Section 2 sets out the basic terminology that will be used throughout and, in particular, the role of the reference technology common to all approaches. Section 3 reviews the non-parametric approach to efficiency estimation originated by Farrell (1957) and developed by Charnes, Copper and Rhodes (1978) into DEA (Data Envelope Analysis). Section 4 reviews the direct index number approach to the calculation of efficiency. Section 5 discusses parametric approaches to efficiency estimation in the context of a static model and section 6 reviews attempts to allow for dynamic structure in the underlying model. Finally, section 7 draws some conclusions from this review.

2. Basic Terminology and Assumptions

Consider a firm which uses N inputs, indexed $n=1,2,\ldots,N$, to produce M outputs, indexed $m=1,2,\ldots,M$. The input bundle is denoted by $x=(x_1,x_2,\ldots,x_N)$ and the input price vector is denoted by $w=(w_1,w_2,\ldots,w_N)$. The output bundle is denoted by $y=(y_1,y_2,\ldots y_M)$ and the output price vector is denoted by $p=(p_1,p_2,\ldots,p_M)$. We assume that all elements of the vectors x,y,w and p are positive (ie. $x,y,w,p>0$). Total cost is given by $C=w.x$.

The production technology of a firm s can be described by an input requirement set, $X^s(y)$. For each output bundle, this set consists of all input bundles from which it can be produced.

We will now make three sets of commonly used assumptions concerning $X^s(y)$. First, and least contentious, is the assumption of regularity. This is composed of the following three separate assumptions:

R1: For any y, there exists an x such that $x \in X^s(y)$.
i.e. for any y there is an x that can produce it.

R2: For any $y>0$, $0 \notin X^s(y)$.
i.e. if any element of y is greater than zero, then at least one element of x must also be greater than zero (the "no free lunch" assumption, or "you can't get something for nothing"!).

R3: For any y, $X^s(y)$ is a closed subset in $\mathbb{R}^n$.

To define efficiency we need a second set of assumptions concerning the disposability of inputs. Here two alternatives can be referred to as weak and strong disposability of inputs.

D1: Weak disposability of inputs requires that if the firm uses proportionally more of all its inputs it may still produce the same outputs. In mathematical notation

$$x \in X^s(y) \Rightarrow \lambda x \in X^s(y) \text{ for any scalar } \lambda \geq 1.$$

D2: Strong disposability of inputs requires that if any input or group of inputs is increased, the firm can still produce the same outputs. In mathematical notation

$$x^+ \geq x \in X^s(y) \Rightarrow x^+ \in X^s(y)$$

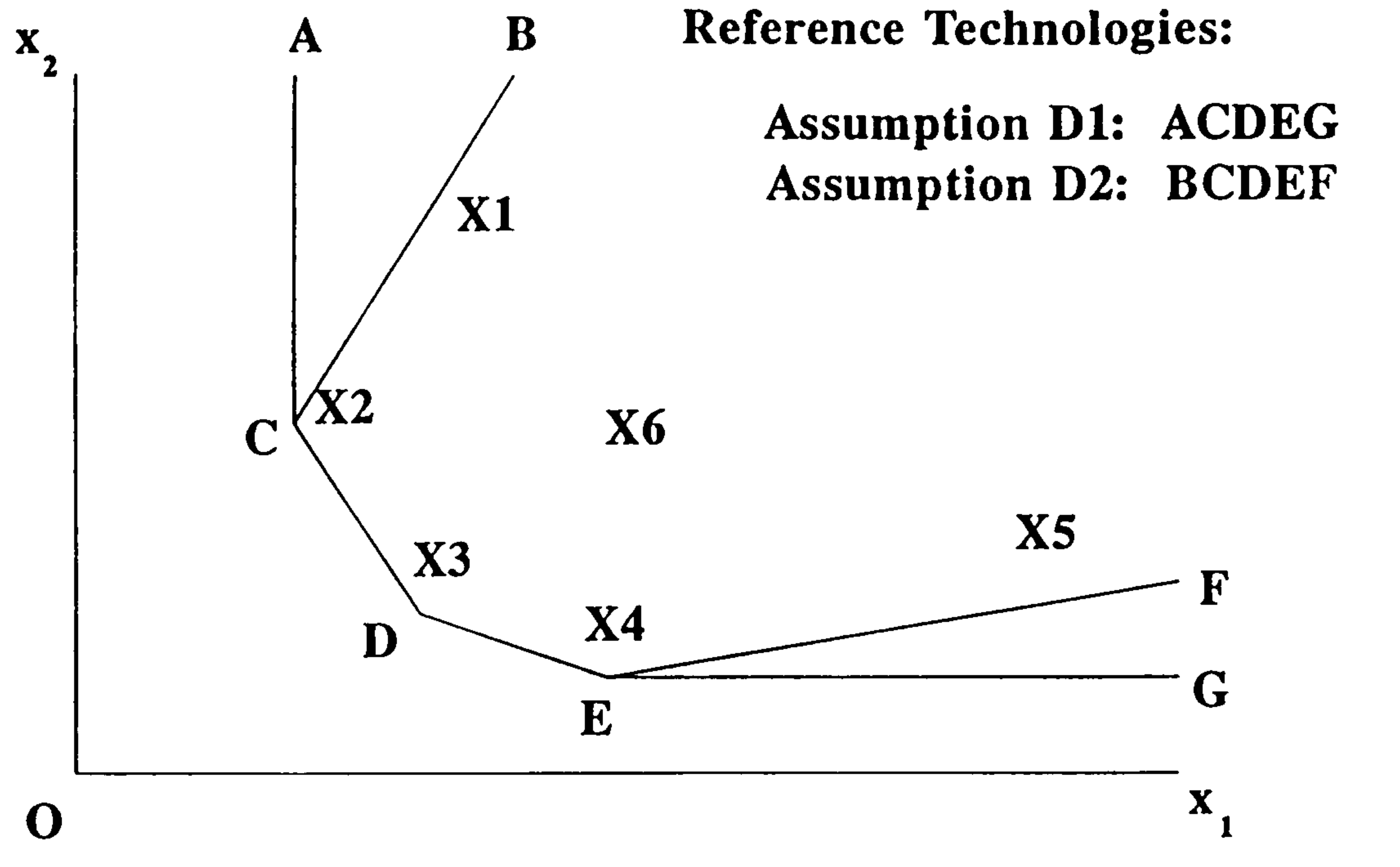
An even stronger assumption concerns the convexity of the input requirement set. This property is sometimes used in efficiency measurement. If x^+ and x^- are both input bundles that can produce y, then any weighted average of x^+ and x^- can also produce y. In mathematical notation, the production technology is convex if and only if

$$x^+ \in X^s(y) \text{ and } x^- \in X^s(y) \Rightarrow \lambda x^+ + (1-\lambda)x^- \in X^s(y) \text{ for any scalar } 0 \leq \lambda \leq 1.$$

Figure I illustrates these assumptions for the case of two inputs, x_1 and x_2. Only weak disposability of inputs (D1) is necessary to define efficiency, although we will also refer to strong disposability of inputs (D2) in our discussion.

Figure I

Disposability of Inputs and Reference Technologies



The disposability of input and convexity assumptions we have just described concern different input combinations that can produce a given output. A final group of assumptions we will refer to cover monotonicity and returns to scale. These concern the input combinations necessary to produce different output levels. These assumptions are, in increasing order of strength, as follows:

M1: Monotonicity can be expressed as

$$X^s(\lambda y) \subseteq X^s(y) \text{ for any scalar } \lambda \geq 1.$$

M2: Non-increasing returns to scale can be expressed as

$$X^s(\lambda y) \subseteq \lambda X^s(y) \text{ for any scalar } \lambda \geq 1.$$

M3: Constant returns to scale can be expressed as

$$X^s(\lambda y) = \lambda X^s(y) \text{ for any scalar } \lambda>0.$$

Figure II illustrates these assumptions for the case of one input and one output, x_1 and y respectively. Although not necessary to define efficiency, we will refer to these assumptions in our discussion of methods of efficiency measurement.

Figure II

Returns to Scale and Reference Technologies

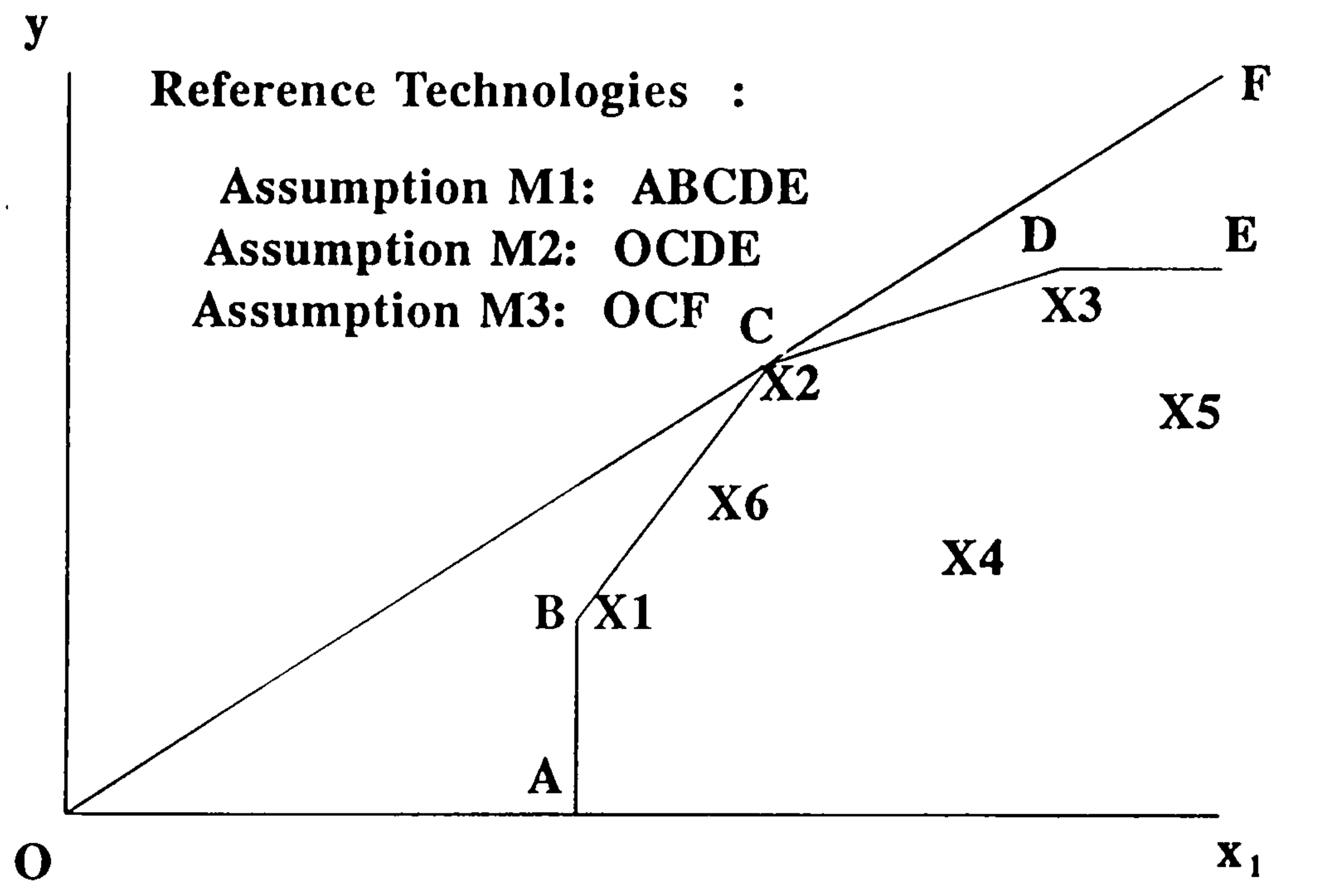


Figure III illustrates an input requirement set in the case of just two inputs (x_1 and x_2), bounded by the curve FF. This input requirement set obeys the assumptions of regularity and disposability of inputs described above, and is also convex.

It is useful to work with the production technology of firm s in terms of its transformation function, $f^s(x,y)=0$. The transformation function can be defined from the input requirement set by the relationship

$X^s(y) \equiv \{ x : f^s(x,y) \geq 0 \}$

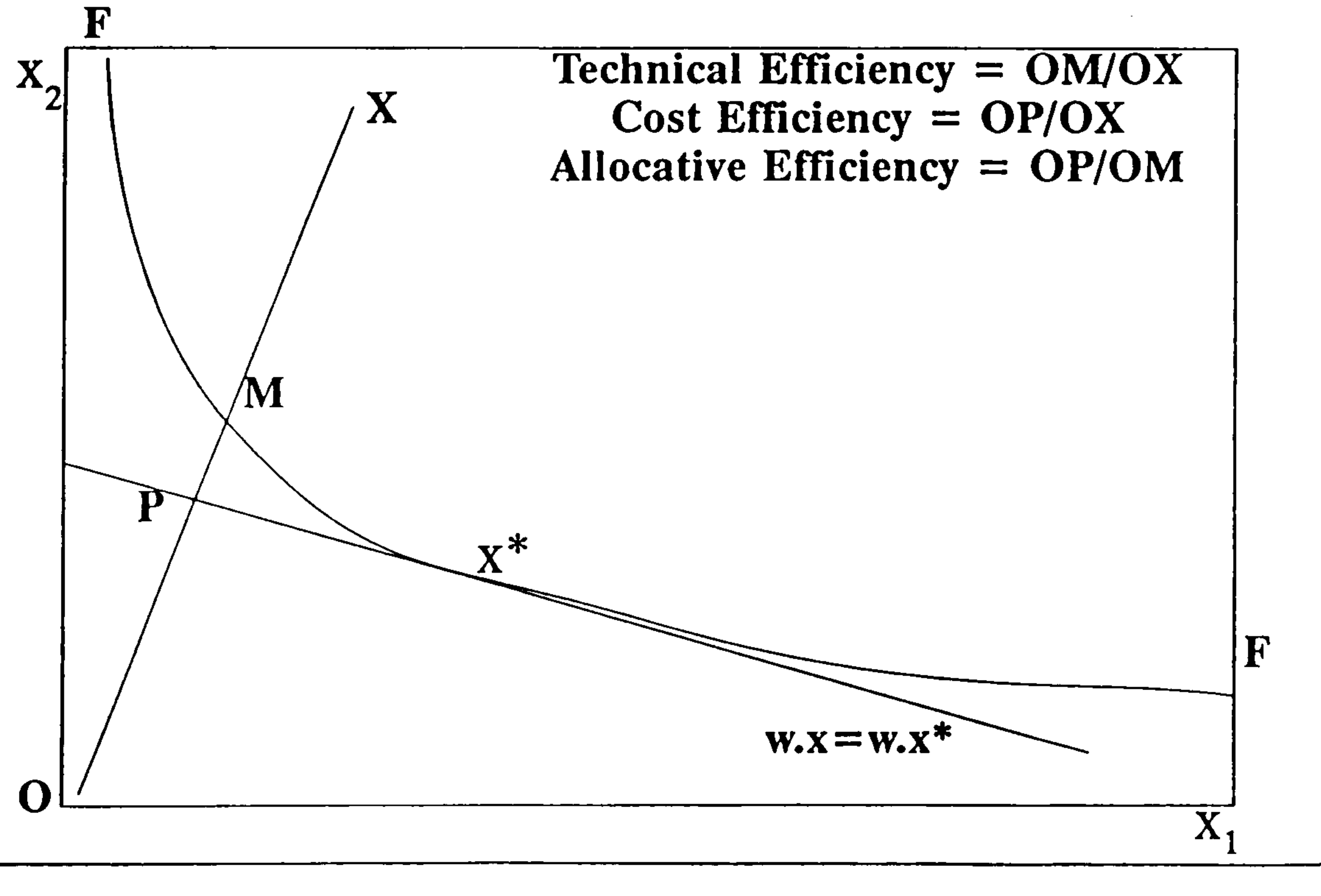
In the case of Figure III where y is fixed, $f^s(x,y)=0$ becomes the curve FF.

Let $x^*(y)$ be the input bundle with which firm s is able to produce y at minimum cost. In mathematical notation x^* is defined by

$w.x^* = \min \{ w.x : x \in X^s(y) \}$

In terms of Figure III, x^* is given by the point X^*, where the hyperplane $w.x^*=w.x$ touches the transformation function FF.

Figure III
Measurement of Efficiency



Consideration of efficiency requires the establishing of a reference technology. We will take the input requirement set $X^s(y)$ to represent the reference technology.

Assuming regularity conditions (R1 to R3) and weak disposability of inputs (D1), we can meaningfully define three types of efficiency as follows:

The input-based Technical Efficiency of producing y with x is defined as

$$P^s(x,y) = 1 \;/\; \max \{ \sigma : (x/\sigma) \in X^s(y) \} \tag{2.1}$$

For σ to be uniquely defined $X^s(y)$ must obey regularity conditions and the weak disposability of inputs assumption (R1 to R3 and D1).

The Cost Efficiency of producing y at cost C when input prices are w is defined as

$$Q^s(C,y,w) = 1 \;/\; \max \{ \sigma : w.x/\sigma \geq w.x^* \} \tag{2.2}$$

Finally, the Allocative Efficiency accounts for the difference between cost efficiency and technical efficiency. That is

$$R^s(x,y,C,w) = Q^s(C,y,w) \;/\; P^s(x,y) \tag{2.3}$$

Figure III illustrates these three measures of efficiency.[4]

---

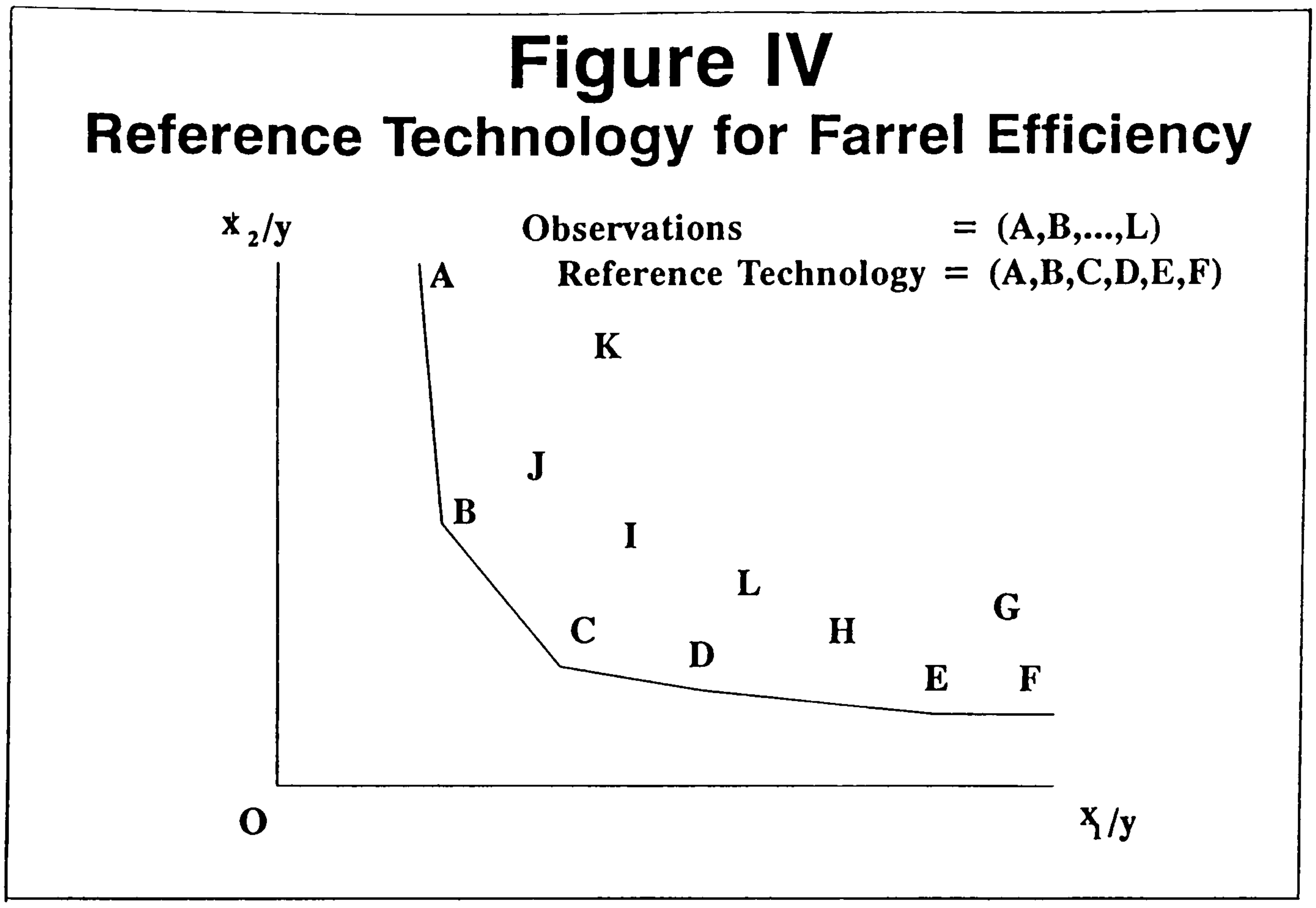
[4] Throughout this thesis we concentrate on input-based efficiency. However, it is also possible to define output-based efficiency (technical, cost or allocative). This simply involves determining the minimum scalar needed to multiply output by, in order to bring the firm onto the reference technology. In the case of constant returns to scale, the two measures will coincide. In all other cases, input-based and output-based efficiency measures will differ. (See Caves, Christensen and Diewert (1982b).)

3. The Farrell Framework

The previous section has discussed the theory of efficiency measurement. Farrell (1957) made a path breaking contribution to empirical efficiency measurement when he proposed forming the reference technology with an efficiency frontier given by fitting a convex hull around the observations of inputs normalised by dividing through by a single output. The convex hull was assumed to obey the regularity assumptions (R1 to R3), weak disposability of inputs (D1) and, in the case of Farrell's original approach, constant returns to scale (M3). This last assumption was necessary in order to allow normalised inputs to be compared at different levels of output.

The approach is illustrated in Figure IV for the case of two inputs (x_1 and x_2) and one output (y). Todd (1985) has applied the Farrell approach to measure the performance of the West German manufacturing industry between 1970 and 1980 using labour and capital as the two inputs.

It should be noted that the precise form of the assumptions for the reference technology will affect the measurement of efficiency. Grosskopf (1986) showed that efficiency measures in a Farrell framework using alternative assumptions about the reference technology are related by some inequalities. In general, stronger assumptions for the reference technology reduce measured efficiency by increasing the distance of observations from the frontier. For example, a reference technology that assumes constant returns to scale (M3) will not result in greater measured efficiency as one that just assumes monotonicity (M1). Similarly, the assumption of strong disposability of inputs (D2) will result in reduced measured technical efficiency than if the reference technology was assumed to obey weak disposability of inputs (D1).



For a long time, Farrell's analysis was restricted to the case of just two inputs and one output, since it relied on graphical methods in two dimensions to fit the convex hull used as the reference technology. A step forward was made by Charnes, Cooper and Rhodes (1978) who showed that when there is only one output the problem can be expressed as a series of linear programmes, and in this way can be numerically solved for any number of inputs. The problem is to find a set of vectors, $\beta^s=(\beta_1^s,\beta_2^s,\dots,\beta_N^s)$, $(s=1,2,\dots,S)$ that

minimises $\quad x^s.\ \beta^s$

such that $\quad x^r.\ \beta^s \geq y^s \quad ; \quad r=1,2,\dots S$

$\qquad\qquad\quad \beta^s \geq 0$

where the superscript $s=1,2,\dots,S$ denotes firms in the sample. The elements of the vector β^s can be interpreted as "prices" for the factor inputs. It will be noted that the Charnes, Cooper and Rhodes model no longer necessarily requires constant returns to scale (M3) (although it can still be imposed as an additional constraint if desired). Essentially, constant returns to scale is replaced by the much weaker assumption of monotonicity (M1).

Furthermore, Charnes, Cooper and Rhodes generalised the problem to many inputs and many outputs by forming a measure of efficiency as the ratio of a weighted combination of outputs to a weighted combination of inputs. Weights are then found to maximise this ratio, subject to the constraint that the ratio (efficiency) be less than or equal to unity for every firm. In mathematical notation, this can be expressed as finding the vectors $\alpha^s=(\alpha_1^s,\alpha_2^s,...,\alpha_N^s)$ and $\beta^s=(\beta_1^s,\beta_2^s,...,\beta_N^s)$, $(s=1,2,...,S)$ that

maximises $\quad y^s.\alpha^s/x^s.\beta^s$

such that $\quad y^r.\alpha^s/x^r.\beta^s \leq 1 \quad ; \quad r=1,2,...S$

$\qquad\qquad\quad \alpha^s,\ \beta^s \geq 0$

in addition to other constraints. As before, the vectors α^s and β^s can be interpreted as "prices" for the outputs and inputs respectively. Charnes and Cooper (1985) formulated a solution based on a non-linear, non-convex and non-Archimedean fractional programming problem.

This approach has given rise to the technique known as Data Envelope Analysis (DEA). Further developments of DEA techniques are discussed in detail in the book by Sengupta (1989).

Empirical applications of DEA are now far too numerous to list in their entirety. Examples of case studies, including some by Charnes and Cooper, can be found in a book collecting together DEA and stochastic frontier studies (discussed below), edited by Fare and Dogramaci (1988). Other examples may be found in the special edition of the Journal of Econometrics (1990).

4. The Direct Index Number Approach

An approach to efficiency measurement that is even older than Farrell's efficiency frontier is the direct calculation of simple factor productivity indices such as labour or capital productivity. These indices can be aggregated into Total Factor Productivity (TFP) (see, for example, Solow (1957)). TFP can be interpreted as measuring shifts in the production function of the firm, industry or economy under study. The use of indices to measure and understand the reasons for productivity growth in individual sectors, and in the economy as a whole, has been practiced by authors such as Jorgenson and Griliches (1967) and Gollop and Jorgenson (1980) in the US and by Peterson (1979) in the UK, and in a series of books by Denison (e.g. Denison (1979) and Denison (1985)) containing very detailed sectorial estimates of TFP.

Theoretical work on the use of indices to measure productivity (or efficiency) has been pioneered by Diewert (1976) and others (chiefly Caves, Christensen and Diewert (1982a and 1982b)).

Samuelson and Swamy (1974) and Diewert (1976) have noted the relationship between underlying economic relationships (such as the production function) and index numbers. All commonly used index numbers are implicitly based on some functional form for the production or cost structure of the firms being compared. For example, fixed weight Laspeyre and Paasche indices are consistent with a Leontief type technology. Diewert (1981) shows the relationships between distance functions (on which efficiency indices are based) and underlying functional forms for utility, production or cost. Conversely, productions functions (or dual cost functions) obeying certain assumptions imply an efficiency index as shown in section 2. Diewert (1976) coined the phrase *exact* to describe index numbers that could be analytically derived from an underlying function that satisfies fundamental economic principles. He also used the phrase *superlative* to describe an index that is *exact* for a *flexible* underlying economic function.

Working from the definitions in section 2, the concern of Diewert (1976) and Caves, Christensen and Diewert (1982a and 1982b) was to find indices that are based on very general classes of production function and would, therefore, be widely applicable. Initially, the work of Diewert (1976) and Caves, Christensen and Diewert (1982a and 1982b) concentrated on the measurement of technical efficiency. The approach these authors adopted was to assume that there is an underlying functional form that describes the technology of both firms. This can be expressed either as a transformation function (such as $f(x,y)=0$) or equivalently as a distance function calculated as the scalar by which all inputs of the firm need to be deflated in order that the firm lies on the underlying transformation function. Effectively, this distance function is the same as the measure of technical efficiency given in section 2 when the underlying transformation function is used as the reference technology. The functional form adopted was the transcendental logarithmic function (or *translog* for short) popularised by Jorgenson and Lau (1977). This function is a *flexible form*, ie. it can provide a second order approximation to any arbitrary function. In its most general form it makes no assumptions on returns to scale or disposability of inputs, although the weak form of the latter (D1) is required in order to unambiguously define efficiency from equations (2.1), (2.2) and (2.3).

Diewert (1976) and Caves, Christensen and Diewert (1982a and 1982b) derived indices to make productivity comparisons directly between firms (or the same firm at different times), either of which could be regarded as the reference. Therefore, for example, a formula for $P^l(x^k,y^k)$ is derived as the technical efficiency of firm k with respect to the reference technology of firm l. Since transitivity can be regarded as a beneficial attribute of index numbers (see Fisher (1922)), Diewert (1976) and Caves, Christensen and Diewert (1982a and 1982b) wished to ensure that both firms were treated symmetrically, so that the indices gave the same result when applied to either firm using the other as a reference. They achieved this by taking a geometric average of the two indices formed when each firm in turn is used as a reference. In mathematical notation the transitive technical efficiency index becomes

$$\log P(x^k,y^l,x^k,y^l) = [\ \log P^l(x^k,y^k) + \log P^k(x^l,y^l)\ ] \ / \ 2 \tag{2.4}$$

where the superscripts k and l indicate the two firms in the comparison.

Caves, Christensen and Diewert (1982a) showed that in the case of constant returns to scale, allocative efficiency and an underlying transcendental logarithmic transformation function, the index in equation (2.4) becomes the Tornqvist index

$$\log P(x^k,x^l,y^k,y^l)$$
$$= \frac{1}{2}\sum_{m=1}^{M} (T^k_m+T^l_m) \log(y^l_m/y^k_m) - \frac{1}{2}\sum_{n=1}^{N} (S^k_n+S^l_n) \log(x^l_n/x^k_n) \tag{2.5}$$

where for firm s:

$T^s_m = p^s_m y^s_m / p^s \cdot y^s$,
the share of total revenue that is earnt by output m;

and $S^s_n = w^s_n x^s_n / w^s x^s$,
the share of total cost that is incurred from input n.

Therefore, the index of equation (2.5) is *superlative*.

Caves, Christensen and Diewert (1982a) extended the result to the cases of both increasing and decreasing returns to scale. Here, they found that an adjustment is required to the revenue share weights used in the index. This more general result is

$$\log P(x^k,x^l,y^k,y^l)$$
$$= \frac{1}{2}\sum_{m=1}^{M} (T^k_m/\varepsilon^k+T^l_m/\varepsilon^l) \log(y^l_i/y^k_i) - \frac{1}{2}\sum_{n=1}^{N} (S^k_n+S^l_n) \log(x^l_n/x^k_n) \tag{2.6}$$

where ε^s is the degree of local returns to scale[5].

The most recent theoretical discovery in this area has been by Fare and Grosskopf (1992), who find that the Malmquist index (on which Diewert (1976) and Caves, Christensen and Diewert (1982a) and (1982b) show the Tornqvist index to be based) can also be used to derive the Fisher ideal index, without the need to make assumptions concerning the coefficients on second order terms of an underlying transcendental logarithmic distance function. (Diewert (1976) and Caves, Christensen and Diewert (1982a) and (1982b) assumed that these coefficients were

---

5 Local returns to scale for firm s is defined as follows: Consider proportionally increasing all inputs, x, by a factor ϕ. Let $u^s(y,x,\phi)$ be a factor of proportionality by which all outputs, y, must be increased so that the inflated input and output vectors lie on the transformation function for firm s; i.e. $u^s(y,x,\phi)$ is the solution to $f^s(\phi x, u^s y)=0$. Differentiating this with respect to ϕ gives

$$x.\frac{\partial f^s}{\partial x} + y.\frac{\partial f^s}{\partial y}\,\frac{\partial u^s}{\partial \phi} = 0$$

$$\frac{\partial u^s}{\partial \phi} = -\,x.\frac{\partial f^s}{\partial x} \Big/ y.\frac{\partial f^s}{\partial y}$$

The degree of local returns to scale is given by

$$\varepsilon^s = \frac{\partial u^s(y,x,\phi)}{\partial \phi} \text{ evaluated at } \phi=1.$$

Therefore, the degree of local returns to scale is given by

$$\varepsilon^s = -\,x.\frac{\partial f^s}{\partial x} \Big/ y.\frac{\partial f^s}{\partial y}$$ evaluated at $\phi=1$ so as to give returns to scale in the locality of (x,y).

If local returns to scale are constant (increasing, decreasing), then $\varepsilon^s=1$ (>1, <1 respectively).

equal between firms.) However, convexity and allocative efficiency are required for this to be true.

The general approach to productivity indices discussed above can easily be applied to cost efficiency (as we do in the next chapter), and disaggregations of the indices into different components of efficiency are also possible (e.g. the effect of individual inputs). This alleviates one of the potential criticisms of the direct index number approach, namely that it provides insufficient information to the investigator. An example of this analysis applied to telecommunications operators can be found in a chapter by Denny and de Fontenay in Fare and Dogramaci (1988).

5. Estimated Parametric Models

Parametric models can be used to estimate a reference technology against which efficiency comparisons can be made. Recently, these models have been fitted using panel data with a residual, u_{st}, associated with firm s in time period t. In fitting models for the purpose of efficiency comparisons, it is important to give careful thought to the assumptions being made about the model residuals since this will have crucial implications for their interpretation in terms of efficiency. Broadly, there are two groups of issues that need to be considered.

First of all there is the question of whether residuals reflect misspecification of the model, or efficiency factors that are modelled through (for example) an error components model. In an extreme view, all efficiencies and inefficiencies of the firm could be regarded as resulting from omitted variables in the underlying structural model. In this case we would seek to explain all variation in terms of structural variables, leaving the residuals with the role of accounting for observation errors alone. When we come to measure efficiency we would then decide which of estimated structural variables were relevant to a measure of efficiency, and exclude the model residuals since they only measure random observation errors. For example, different labour practices may have an impact on costs of production and could be incorporated as a variable in the structural model of the firm. However, to the extent that they are under the control of the firm, we would want to include their effect in our measurement of efficiency.

At the other extreme, our model of the firm may be a reduced form of exogenous variables alone. In the case of a cost function, we would only include output and factor prices (assuming that these are genuinely exogenous). In this case the model residuals would consist of much more than just observation errors. They could also include the effects of many other management and production decisions that affected the performance of the firm, and that we would wish to include in our measurement of efficiency (for instance, the decisions on what labour practices to adopt). In this case, efficiency would be measured by the size of the model residuals.

We will return to this issue in later chapters when we discuss adjustment costs. I will argue that "optimal"[6] adjustment costs are best included in the structural model of the firm, and excluded from the measurement of efficiency, ie. if a firm suffers adjustment costs in raising output, the associated minimum increase in costs should not be counted as inefficiency. Nevertheless, the importance of adjustment costs to the cost structure of the firm is of interest in its own right.

The second group of issues that need to be considered concern the statistical model for the behaviour of the residuals themselves. Our view of how these residuals behave will affect the absolute measure of efficiency for each firm, and also the relative efficiency estimates to the extent that they alter the estimate of the underlying reference technology. The choices of assumption for the residuals can be divided

---

6 That is, the adjustment costs associated with the minimum total cost (the usual factor payments plus adjustment costs) necessary to achieve a given change in the levels of outputs.

into two broad types.

First, the residuals (depending on sign) can be thought of as representing efficiency and inefficiency factors for the observed firms (and are usually assumed to be normally distributed). The equation can be estimated by a least squares or maximum likelihood technique and the resulting model will represent a reference technology based on the *average* performance of the sample of firms. Examples of this approach are Berndt and Khaled (1979) who include a general discussion of appropriate parametric specifications and an application to the US manufacturing sector, Anandalingam and Kulatilaka (1987), who use their parametric specifications to decompose efficiency into technical, allocative and structural components, and Callen (1988) with an application to a sample of US electricity generating companies in which he investigates the status of capital as a variable or quasi-fixed input in the parametric specification.

Second, a parametric efficiency frontier could be fitted by assuming that the residuals follow a one-sided distribution (such as the gamma distribution). The reference technology can be thought of as representing the best practice of the sample of firms. Forsund, Lovell and Schmidt (1980) give a good introduction to these kinds of models. The simplest estimator is COLS (corrected ordinary least squares). Here, a cost or production function is first estimated by OLS to give best linear unbiased estimates of all the slope coefficients in the regression. Then the constant in the estimated equation is adjusted by the minimum amount necessary to ensure that all observations fall either on or above (in the case of a cost function), or on or below (in the case of a production function) the estimated frontier. More efficient procedures that allow direct estimation of the frontier employ maximum likelihood. Greene (1980a) found sufficient conditions on the one-sided distribution of the residuals for maximum likelihood to yield consistent and asymptotically efficient estimates. The gamma distribution, for example, satisfies these conditions. Greene (1980b) gives an empirical example of the use of the maximum likelihood estimator in this type of model.

Instead of attributing the whole of the residual to efficiency related causes, the model residuals can be assumed to be composed of two additive components: one following a one-sided distribution such as the gamma (representing inefficiency) and the other following a two-sided distribution such as the normal (representing model error). We will refer to these components as v_{st} and w_{st} respectively, so that $u_{st}=v_{st}+w_{st}$. This type of assumption would be consistent with a reference technology representing best practice, and the possibility of measurement or other types of error. If ignored, such errors would contaminate estimation of the frontier. These models are usually referred to as *stochastic frontiers.*

It should be noted that in the most general stochastic frontier model, inefficiencies cannot be identified for each firm in each individual time period (since there is no way of determining how much of each individual residual relates to inefficiency and how much to model error). The most usual assumption is that inefficiencies vary between firms but are fixed over time (so that $v_{st}=v_s$). This allows the identification of firm specific inefficiencies.

The simplest estimator for a stochastic frontier is again the COLS estimator. However, the procedure differs from that used to estimate a deterministic frontier in that the adjustment to the equation's constant is based on an estimation of the parameters of the assumed underlying one-sided distribution (usually a gamma distribution) using the moments of the OLS residuals. More efficient direct estimators for these types of models were first proposed by Meeusen and van den Broeck (1977) and Aigner, Lovell and Schmidt (1977). These early attempts assumed distributions for v_{st} with modes at $v_{st}=0$ (such as the half-normal or negative exponential distribution). Unfortunately, Greene (1980a) showed that these distributions will not satisfy maximum likelihood regularity conditions. A gamma distribution does, however, satisfy these conditions. Stevenson (1980) proposed two maximum likelihood estimators assuming non-zero moded distributions. These were a truncated normal distribution and a gamma distribution (which he notes can be considered as a generalisation of the negative exponential). Using two empirical datasets he applied his truncated normal distribution model and rejected the hypothesis of an inefficiency mode at zero for both datasets.

Sickles and Streitwieser (1989) have an application of the stochastic frontier technique to the US gas pipeline industry in which they also incorporate autoregressive errors. In this paper, the one-sided residuals representing inefficiencies are taken to be firm-specific but constant throughout the sample period. This allows the estimation of firm inefficiencies. The two-sided residuals can take a firm-specific autoregressive structure that controls for firm-specific effects that may contaminate the error structure. In mathematical notation

$$v_{st} = v_s \quad \text{(for all } t\text{)}$$

$$w_{st} = \rho_s w_{st-1} + \zeta_{st}$$

where ρ_s is a firm specific autoregressive coefficient and ζ_{st} is a white noise residual. Cornwell, Schmidt and Sickles (1989) develop the statistical estimation and inference theory for this type of model with the addition that firm-specific inefficiencies are allowed to vary over time according to a quadratic time trend.

Kumbhakar (1991a) summarise the maximum likelihood estimators for a case where both fixed firm-specific and time-specific effects are drawn from two-sided distributions, and inefficiencies associated with each observation are drawn from a one-sided distribution. That is

$$u_{st} = w_{1s} + w_{2t} + v_{st} + \zeta_{st}$$

$$\begin{aligned} \text{where } w_{1s} &\sim \text{iid } N(0,\sigma_1^2) \\ w_{2t} &\sim \text{iid } N(0,\sigma_2^2) \\ v_{st} &\sim \text{iid } N(0,\sigma_3^2) \quad ; \quad v_{st} \le 0 \\ \zeta_{st} &\sim \text{iid } N(\zeta,\sigma_4^2) \end{aligned}$$

In a further paper, Kumbhakar (1991b) develops an algorithm to decompose the observed inefficiencies from a stochastic frontier transcendental logarithmic cost function into technical and allocative components. The latter are derived from the residuals of the estimated factor share equations, developing a method of Schmidt (1984) in which allocative inefficiencies are modelled as a quadratic form of the factor share residuals. Kumbhakar (1991b) imposes an additional reasonable constraint on Schmidt's method by requiring that factor shares with allocative inefficiency integrate back up to the original transcendental logarithmic cost function. We refer to this approach in chapter IV when we want to estimate the impact of adjustment costs (which in a static model appear as allocative inefficiencies) on total costs.

The choice between different estimators for stochastic frontier models is considered in a Monte Carlo study by Gong and Sickles (1992). They consider three econometric estimators (in addition to DEA) for a model in which technical efficiency is assumed to be fixed (at v_s) for each firm over time. They model inefficiency as the distance from a production frontier or cost frontier dual. The three estimators are maximum likelihood, generalised least squares (using a transformation to re-write the model so that the residuals for each firm have zero mean and constant variance) and a *within* estimator in which the model is expressed as deviations over time from individual firm means. This latter estimator avoids two problems of stochastic frontiers applied to technical efficiency. First, correlation between inputs and technical efficiency itself may bias inefficiency estimates. However, the *within* estimator removes this correlation before model estimation. Second, is the problem of the dependence of the results on the distributional assumption concerning the form of technical efficiency (the distribution of v_s). Again the *within* estimator avoids this problem by removing v_s before model estimation. Gong and Sickles conclude by preferring this *within* estimator.

Examples of these kinds of models are found in the special edition of the Journal of Econometrics (1990) on efficiency frontiers. Of particular relevance to this thesis is a case study by Ferrier and Lovell (1990) dealing with the cost efficiency of the banking industry. We discuss the results of this paper in more detail in later chapters. Now, however, we note that Ferrier and Lovell found a large degree of allocative inefficiency in their study, especially when using the parametric model approach. This finding is also supported by work of Drake and Weyman-Jones (1992). They analyse a sample of UK building societies, finding that large inefficiencies result from the societies choosing the wrong combinations of factor inputs. It seems possible that this may be in part caused by un-specified dynamic effects such as a failure of firms to immediately adjust factor inputs to the static *optima* (following changes in outputs or factor input prices). These effects might be explained by inclusion of adjustment costs in the model. The modelling of such dynamic effects (especially in the banking sector) is a central theme in this thesis.

Further stochastic frontier case studies are to be found in the book edited by Fare and Dogramaci (1988), already cited. In particular, this volume contains further contributions by Lovell and Schmidt.

One of the main concerns with stochastic frontier models is the need to impose an additional assumption on the inefficiencies (v_{st}) in order to distinguish them from the usual model residuals (w_{st}). The most common assumption is that inefficiencies are fixed over time for each firm ($v_{st}=v_s$). This may be unrealistic, particularly when a medium length time series is being considered, and especially in a competitive market where persistently inefficient firms would fail to survive.

A second concern is that in order to produce efficient maximum likelihood estimates, a particular distribution must be assumed for the inefficiencies. There are no good *a priori* reasons for preferring one one-sided distribution over another, save the statistical convenience of Greene's result (1980a). This is that in order to ensure the estimates possess desirable asymptotic properties the density of v_{st} should be zero at $v_{st}=0$ and the derivative of the density of v_{st} with respect to its parameters should approach zero as v_{st} approaches zero. Although the gamma distribution conveniently satisfies these criteria, there is no other theoretical reason for adopting this distribution.

A third concern is that it is not necessarily clear that a stochastic frontier estimation in which a composed residual of, say, a gamma distribution and a normal distribution, is actually identifying inefficiency and measurement error respectively. This critically depends on the assumption that measurement errors are normally distributed and inefficiencies are gamma distributed. However, it is not impossible to imagine situations in which measurement errors are gamma distributed and inefficiencies are normally distributed (or drawn from a truncated normal distribution). In this case, inefficiency estimates will be biased.

6. Dynamic Models

So far all the approaches we have discussed for measuring efficiency have been based on single period models. The only attempts to take account of dynamic effects in either the production or cost structure of the firms have been in the context of parametric models (although we do note that Bartelsman and Dhrymes (1991) have attempted to use transition matrices to track the persistence of firm specific productivity between time periods using both TFP and estimated production functions).

The first substantial attempt to introduce theoretically consistent dynamic effects into an estimatable cost function was that by Morrison and Berndt (1981). These authors built on previous work by Treadway (1974) and Berndt, Fuss and Waverman (1977 and 1980). Their approach was to take some factor inputs to be quasi-fixed and subject to increasing adjustment costs so that as purchases increase in any one period an amount of foregone output arises. This implies a production function constraint which is augmented to be $f(x,z,\dot{z},y)=0$ where z is a vector containing the quasi-fixed inputs. Taking z and $\dot{z}$ to be fixed, Morrison and Berndt then solved the short-run problem of choosing the variable factor inputs to minimise the expected present value of the cost of producing a given flow of output, subject to the production function constraint, $f(x,z,\dot{z},y)=0$. In writing the expected present value of future costs they assumed factor prices are expected to rise at a constant exponential rate. They also assumed static expectations for future output levels. A short-run solution to this problem can then be written as $g(w,z,\dot{z},y)=C$ where g is a function obeying certain regularity assumptions.

The long-run solution in which the optimal quasi-fixed factor inputs (z^*) are solved can be implicitly found from the Euler first-order conditions. Treadway (1974) showed that this also corresponds to an approximate solution to the flexible accelerator model.

In order to implement the model empirically, a functional form must be specified for $g(w,z,\dot{z},y)$. Morrison and Berndt (1981) assumed a quadratic cost function restricted to impose long-run constant returns to scale. Factor demand equations (normalised by output) were derived and estimated as part of the system for both variable and quasi-fixed factor inputs.

The same authors have recently conducted a further case study (Morrison and Berndt (1991)). Again using a parametric model, the authors made inferences about the productivity of information technology equipment in US manufacturing industries in three ways. First, they calculated the ratio of the shadow value of information technology equipment ($-\partial G/\partial x$, where $G(.)$ is a variable cost function inclusive of adjustment costs, and x is a vector of factor inputs) to the ex ante rental price. This ratio shows that in 1986 (the last year of their data sample) the estimated marginal benefits of investment in information technology (the shadow value) were, in general, less than the rental prices - implying over-investment. Second, the authors concluded that the estimated elasticities of demand for labour with respect to changes in the stock of information technology equipment was increasing in absolute magnitude over time. Finally, the authors concluded that the estimated elasticity of technical progress with respect to the stock of information technology equipment was very small in magnitude, indicating that increases in such equipment have only a small impact on technical progress. The richness of these conclusions illustrates the main benefit of a parametric approach to productivity measurement. However, the authors reduced form for the variable cost function, $G(.)$, does not allow them to make an explicit estimate of adjustment costs along the lines that we will attempt in chapter IV.

A further empirical application of the Berndt, Fuss and Waverman (1977 and 1980) approach can be found in Lawrence (1990). This author uses essentially the same model applied to export supply and import demand data, taking capital to be a quasi-fixed input. His primary objective was to estimate short and long run export-supply and import-demand elasticities.

In this section we have restricted our discussion to dynamic models which have been used as a basis for measuring efficiency. However, other authors have used similar models to analyse other aspects of the theory of the firm. For example, Nickell (1986) uses a dynamic model of factor demands to look at the demand for labour. In fact, the modelling approach we adopt in Chapter IV to derive factor demands has more similarities with that of Nickell (1986) than Morrison and Berndt (1981).

7. Conclusions

It will be helpful to conclude this chapter by reviewing the relative advantages and disadvantages of the different approaches to efficiency measurement discussed. For example, Jorgenson and Griliches (1967) demonstrate the biases in productivity indices that can result from incorrectly separating the change in value of new investment goods into volume and price change components (in the case of the US economy). However, the same problems can arise with all the other approaches described in this chapter.

Some papers and volumes have already attempted evaluation of two of the approaches - DEA and stochastic frontier regression. A book edited by Fare and Dogramaci (1988) collects together papers dealing with both approaches from a very wide range of applications (covering agriculture, manufacturing, telecommunications, airlines and automobiles). One of these contributions is a paper by Banker, Charnes, Cooper and Maindiratta (1988). They compare the performance of DEA with a parametric model based technique through a simulation study. They claim a superiority for DEA, even when the simulated production structure takes the same form as that of the estimated parametric model. The only weakness they find in DEA is in dealing with "corner" observations with a very small or very large quantity of at least one of the inputs or the outputs. However, Banker, Charnes, Cooper and Maindiratta chose a very particular distribution for the efficiency factors. 30% of observations were assumed to be efficient, whilst the remaining 70% were assumed to be drawn from a uniform distribution. Clearly, conventional regression estimates would not be statistically efficient under these circumstances. A paper by Ferrier and Lovell (1990) also compared results from these two techniques applied to data drawn from the banking industry. They conclude that the relative disadvantages of the two techniques can be described as follows. Whereas DEA ascribes noise in the data to inefficiency, parametric approaches risk ascribing error from misspecification of the model to inefficiency.

Even more recently, Gong and Sickles (1992) conducted a Monte Carlo study to compare the performance of DEA and parametric stochastic frontier models. They concluded that the choice between the two methodologies depends critically on how well the underlying technology or cost structure (e.g. the production function in the case of technical efficiency) is captured in the functional form assumed by the investigator. If the assumed functional form used in a parametric model is close to that which exists in reality, then stochastic frontier models out-perform DEA. However, as misspecification of the functional form becomes more serious, the relative performance of DEA increases. DEA also has strong advantages when regressors are thought to be correlated with the inefficiencies being estimated.

In this conclusion, we present our own evaluation of the various classes of methods available to the investigator, including direct index number approaches (which were not considered by most of the papers and collections cited above, although Bartelsman and Dhrymes (1991) do compare direct TFP indices and productivity estimated Cobb-Douuglas and transcendental logarithmic production functions).

a. *Farrell Methods*

Farrell methods (and more recently DEA) appear to have the advantage of not requiring the investigator to pre-suppose any explicit functional form on either the production function or cost function reference (with consequent avoidance of errors in efficiency estimation if this functional form is inappropriate). Constraints are imposed to ensure that the production or cost structure obeys fundamental assumptions of production theory (such as, for example, disposability of inputs) in addition to convexity or concavity respectively. Furthermore these methods do not assume any particular statistical distribution for the underlying data generating model.

However, Farrell methods are not entirely independent of assumptions on reference technologies since, as Grosskopf (1986) showed, different assumptions on disposability and returns to scale which may be imposed on the efficiency frontier fitted around the observations may result in different efficiency estimates.

Set against this is the disadvantage that the reference technology is often based on only a small proportion of the observations that define the efficiency frontier. Errors in these observations will be reflected in the efficiency estimates. Correspondingly, most observations will have no influence on the reference technology (within bounds). However, the sensitivity of Farrell methods to errors in observations on or near the frontier can be used to detect such errors. This will only be possible if a close review is made of all observations forming the efficiency frontier.

A further drawback of Farrell methods is that no statistical inference techniques will be available to test conclusions. Although some attempts have been made to incorporate statistical models into DEA (e.g. Timmer (1971)), they have not been altogether successful, often imposing restrictions on the DEA model (such as assuming the frontier to be based around just one hyperplane rather than a convex hull).

Many studies that have compared DEA and stochastic frontier approaches to inefficiency measurement find that stochastic frontiers attribute a very large proportion of the residuals to observation or model error rather than inefficiency (Aigner, Lovell and Schmidt (1977) and Drake and Weyman-Jones (1992)). By construction, negative observation or model errors on the cost function (or positive errors on the production function) which are either on the frontier itself or larger than the inefficiency of that observation, will shift the Farrell efficiency frontier down (or up in the case of a production frontier). Where observations originally lie on the frontier, the magnitude of this shift will be equal to the magnitude of the error. Conversely, positive observation or model errors on the cost function (or negative errors on the production function) will only affect the Farrell efficiency frontier if they happen to actually lie on the frontier itself, in which case they will shift it up (or down for a production frontier). In this case the magnitude of the shift will never be greater than the error (since the observation may cease to be part of the frontier). In either case (positive or negative errors) there will always be an incorrect estimate of inefficiency for that observation. Therefore, not only may Farrell based methods (that take no formal account of observation or model error) attribute observation or model error to inefficiencies, but in certain cases the whole efficiency frontier may be shifted, most often in the direction of over-stating inefficiencies.

b. *Direct Indices*

The initial attractions of a direct index number approach to efficiency measurement is that index numbers are generally easy to compute and have modest data requirements. Furthermore, they require no estimation to obtain a reference technology.

Nevertheless, all commonly used indices assume an underlying implicit functional form for the technology or cost structure of each observation and reference base. For example, fixed weight Laspeyre and Paasche indices are consistent with a Leontief type technology. Caves, Christensen and Diewert (1982a and 1982b) showed that if the Tornqvist index is to yield an exact measurement of efficiency, then cost minimising behaviour must be assumed and the transformation or cost function is required to have a flexible transcendental logarithmic specification. The assumption of cost minimisation is necessary to derive cost share weights for inputs and implies allocative efficiency. This means that static indices are unsuitable when, for example, adjustment costs result in firms choosing allocatively inefficient factor input mixes within individual time periods. The index number approach we develop in chapter III allows for adjustment costs and consequently results in an index with "adjusted" weights.

The substantial drawback of the index number approach is that it does not estimate the underlying production or cost structure, even though these are implicitly assumed in the formulation of the index. Because the underlying technology or cost function is only implicitly assumed, and not explicitly estimated, there is no guidance on the choice of which outputs and inputs are appropriate, which, if any, inputs should be regarded as quasi-fixed, and whether the implicit functional form of the transformation or cost function is correct.

c. *Parametric Methods*

The relative advantages and disadvantages of parametric models correspond to the disadvantages and advantages of the non-parametric methods. The reference technology is restricted to a particular functional form (which can be fairly general), and certain characteristics of the distribution of the inefficiencies will need to be specified (e.g. a one or two-sided distribution, fixed for each firm over time, etc.). However, during the modelling process the investigator will be able to test the appropriateness of this functional form and will also glean considerable insights to assist in interpreting the resulting efficiency estimates. Furthermore, formal statistical tests will be available for assessing the significance of efficiency differences between firms or time periods.

d. *The Approaches Adopted in this Thesis*

In this thesis we will explore two of the above methods: direct index numbers and parametric models and, in particular, consider their extension to take into account adjustment costs. Although, in general, all the approaches considered in this chapter have their own distinctive advantages and disadvantages, the two we have selected to apply appear to be the most suited for adaptation to the type of dynamic model we intend to use. In chapters III and IV it will be seen that cost functions that model the intricacies of adjustment cost processes are required to have complicated non-linear functional forms. Constraints that re-produce these non-linearities in mathematical programming problems would be impractical, leaving direct index numbers and parametric model approaches as the only feasible alternatives.

CHAPTER III

EFFICIENCY INDICES IN THE CONTEXT OF DYNAMIC COST STRUCTURES

ABSTRACT

This chapter begins by showing how the Tornqvist cost efficiency index can be derived from a transcendental logarithmic cost function, and goes on to extend the result to cost structures that incorporate adjustment costs. Finally, the chapter ends with an empirical application to the UK banking and building society sectors.

1. Introduction

Traditionally, efficiency indices are presented in the context of a static model of the firm. In this chapter we investigate how these indices may be adapted to provide an assessment of the performance of firms when the underlying cost structure includes dynamic adjustment cost effects. We specify that these dynamic effects result from the firm incurring additional costs in any time period in which it changes its use of factor inputs. The cost minimising firm will, therefore, take account of inputs in the previous time period and will not necessarily employ the same mix of factors as a firm with no adjustment costs. Under a static measure of efficiency, this may be interpreted as an allocative inefficiency, where in fact it is optimal behaviour for the cost minimising firm. Static efficiency indices are one period "snapshot" measures and are unable to allow for the effect of adjustment costs.

Diewert (1976) and Caves, Christensen and Diewert (1982a and 1982b) have shown how the Tornqvist Index can be used to measure the efficiency of a firm as a ratio of aggregated output volume to aggregated input volume, under the assumption of a transcendental logarithmic production structure. The principal objective of this chapter is to study the impact on efficiency indices of dynamic adjustment costs within the cost structures of the firm being studied. In particular, we are interested in investigating how static efficiency measures are affected by the presence of adjustment costs and, consequently, how these efficiency measures may be modified to reflect the impact of adjustment costs on the behaviour of cost minimising firms.

The chapter will proceed with four further sections. Section two will give a detailed statement of the assumptions we are making concerning the technology available to the firms being analysed. We will also state their behavioural objectives, which can be followed through to statements about the mix of inputs that they will aim to achieve. This far we will be following the usual analysis of the firm that may be found in standard technical texts such as Varian (1978) or more advanced texts such as Fuss and McFadden (1978). Section two will also introduce two types of efficiency and show how they can be measured by indices in the context of the usual static model. Section three will introduce dynamic features into the model. In particular, we will specify adjustment costs that are a function of the current and previous period's level of factor inputs. We will then investigate how efficiency indices may be appropriately adapted. Finally, section four will present an empirical application to the UK banking and building society sectors.

2. Efficiency Measurement in Static Production Models

Consider a firm which uses N inputs indexed $n=1,2,...,N$ to produce M outputs indexed $m=1,2,...,M$. The input bundle is denoted

$x=(x_1,x_2,\dots,x_N)$ and the input price vector is denoted $w=(w_1,w_2,\dots,w_M)$. The output bundle is denoted $y=(y_1,y_2,\dots,y_M)$ and the output price vector is denoted $p=(p_1,p_2,\dots,p_M)$. We assume that all elements of the vectors x,y and w are positive (ie. $x,y,w,p>0$). We will require that the firm s possesses a differentiable transformation function, $f^s(x,y)=0$, which we take to be increasing and convex in x, and declining and concave in y. This function defines an input requirement set for the firm which we shall call $X^s(y)$.

We assume that an efficient firm will restrict its choice of input and output combinations to those that achieve the following criteria:

(i) Given an output bundle, the firm chooses an input bundle combination so as to minimise costs.

(ii) Given an input bundle, the firm chooses an output combination so as to maximise revenue.

These two objectives ensure that the firm makes best use of its inputs and produces the best combination of outputs without necessarily assuming that the firm maximises profits.

The minimisation problem in (i) can be re-written so as to define the cost function for firm s

$$c^s(y,w) = \min \{ \ w.x \ : \ x \in X^s(y) \ \} \qquad (3.1)$$

It is well known that $c^s(y,w)$ is homogeneous of degree one in w, continuous, non-decreasing and concave in w, and non-decreasing in y (see, for example, Varian (1978) and Fuss and McFadden (1978)). We assume $c^s(y,w)$ is differentiable.

We are interested in measuring how far firms diverge from their production frontier. The most natural way to do this is through a distance function defined as

$$d^s(x,y) = \max_\sigma \{ \ \sigma \ : \ f^s(x/\sigma,y) \geq 0 \ \} \qquad (3.2)$$

Thus $d^s(x,y)$ is the scalar by which we have to deflate each input in order for it to lie on the production frontier of firm s. Such a scalar can always be found.

The regularity and other conditions that apply to the transformation function, $f^s(x,y)$, can be shown to determine those that apply to the distance function $d^s(x,y)$. In particular $d^s(x,y)$ will be monotonic and convex in x. (See Diewert (1981)).

We can now define a calibration of technical efficiency in terms of the distance function. This is done in relation to two firms, $s=k$ and $s=l$ (or alternatively one firm at different times). The index of technical efficiency of firm l relative to firm k using firm k's production technology is defined by the ratio

$$P^k(x^k,x^l,y^k,y^l) = d^k(x^k,y^k) \;/\; d^k(x^l,y^l) \tag{3.3}$$

which is a pure number. Note that if firm k is technically efficient relative to its own production technology, then $d^k(x^k,y^k)=1$.

Similarly, the index of technical efficiency of firm l relative to firm k using firm l's production technology is given by the ratio

$$P^l(x^k,x^l,y^k,y^l) = d^l(x^k,y^k) \;/\; d^l(x^l,y^l) \tag{3.4}$$

We now have two indices to compare the technical efficiency of firms k and l - one from the point of view of firm k, and the other from the point of view of firm l. It would be desirable for the chosen index to be symmetrical so that the same efficiency was calculated for firm l relative to firm k irrespective of whether the base technology used was that of firm l or firm k. ie.

$$P^k(x^k,x^l,y^k,y^l) = P^l(x^k,x^l,y^k,y^l) \tag{3.5}$$

(see Fisher (1922)). In general, this will not be the case. The obvious compromise that remedies this is to take the geometric mean of the two indices, thus

$$\log P(x^k, x^l, y^k, y^l)$$
$$= \{\log[d^k(x^k,y^k)/d^k(x^l,y^l)] + \log[d^l(x^k,y^k)/d^l(x^l,y^l)]\} \;/\; 2 \qquad (3.6)$$

It will soon be seen that this measure of productive efficiency has additional advantages when applied to a particular general class of transformation functions.

It may be of interest to digress at this stage and note that an alternative measure of efficiency can be obtained from finding the maximum scalar by which all outputs of firm l may be multiplied beyond a level y whilst keeping the inputs necessary to produce the inflated output within the input requirement set $X^k(y)$. This is known as "output-based efficiency", and under constant returns to scale will equate to the "input-based efficiency" discussed above (see Caves, Christensen and Diewert (1982a and 1982b)).

Returning to our case of "input-based efficiency", Caves, Christensen and Diewert (1982) make further progress by assuming that $\log d^s(x,y)$ is a transcendental logarithmic function. Such functions were first suggested by Jorgenson and Lau (1977) as providing a second order approximation to any arbitrary function, and, as such have a flexible form. If the distance function, $\log d^s(x,y)$, is a transcendental logarithmic function, the production function of firm s will also be transcendental logarithmic. Empirical studies employing transcendental logarithmic production and cost functions are numerous. For example, see Berndt and Wood (1979), Nissim (1982), Robertson, Caves, Christensen and Tretheway (1984) and Callan (1988).

Caves, Christensen and Diewert (1982) make the further assumption that the coefficient on the second order terms in $\log d^s(x,y)$ are the same for all firms (ie. independent of s). If firms k and l are both producing on their production frontiers so that $d^k(x^k,y^k)=d^l(x^l,y^l)=1$ (ie. both are efficient relative to their own technology) and choose their factor input mix to minimise costs (ie. are allocatively efficient), then Caves, Christensen and Diewert (1982) in their Theorem 4 (page 1407) show that the technical efficiency index between the two firms can be shown to be

$$\log P(x^k,x^l,y^k,y^l)$$
$$= \frac{1}{2}\sum_{m=1}^{M}\left(T_m^k/\varepsilon^k+T_m^l/\varepsilon^l\right)\log\left(y_m^l/y_m^k\right) - \frac{1}{2}\sum_{n=1}^{N}\left(S_n^k+S_n^l\right)\log\left(x_n^l/x_n^k\right) \tag{3.7}$$

where for firm s:

$T_m^s = p_m^s y_m^s / p^s.y^s$,
the share of total revenue that is earnt from output m;
$S_n^s = w_n^s x_n^s / w^s.x^s$,
the share of total cost that is incurred by input n;
and ε^s is the degree of local returns to scale[7].

This is almost a Tornqvist Efficiency Index, except for an adjustment for local returns to scale to the revenue share of each output, reducing exactly to the Tornqvist Index in the case of $\varepsilon^k=\varepsilon^l=1$ (constant returns to scale in both firms).

So far we have discussed a narrow definition of technical efficiency. We now move on to a broader definition of cost efficiency where we incorporate an assessment of the extent to which the firm is producing its output bundle at minimum cost.

As with technical efficiency, we require a means of calibrating cost efficiency between firms (or time periods). The most natural way to do this is with a distance function defined as

$$e^s(C,y,w) = \max_{\sigma} \{ \sigma : c^s(y,w) \le C/\sigma \} \tag{3.8}$$

where $C=w.x$ is the actual observed cost incurred by the firm, and $c^s(y,w)$ is the minimum achievable cost (as already defined by equation (3.1)).

Thus $e^s(C,y,w)$ is the maximum scalar by which we must deflate total

7 We defined the degree of local returns to scale in chapter 2.

cost in order that it lies on the cost function of firm s. $e^s(C,y,w,)$ will be homogeneous of degree one in w, non-increasing and convex in w, and non-increasing in y. $e^s(C,y,w)$ will also be differentiable.

We can now define a calibration of cost efficiency in terms of the distance function. The index of cost efficiency of firm l relative to firm k using firm k's cost structure is defined by the ratio

$$Q^k(C^k,C^l,y^k,y^l,w^k,w^l) = e^k(C^k,y^k,w^k) \;/\; e^k(C^l,y^l,w^l) \tag{3.9}$$

which is a real number. Note that if firm k is cost efficient relative to its own production technology and cost structure, then $e^k(C^k,y^k,w^k)=1$.

Intuitively, this same argument can be developed as follows. Suppose firm l is using an input bundle x^l to produce an output bundle y^l. We say that firm l is cost efficient relative to its own production and cost structure if and only if x^l is such that $w^l.x^l=c^l(y^l,w^l)$. Otherwise, x^l must lie elsewhere within the firm l input requirement set and firm l can be said to be inefficient relative to its own production technology and cost structure.

We can also define firm l cost efficiency relative to firm k's cost structure. We can say that firm l is over-efficient relative to firm k if and only if $w^l.x^l\leq c^k(y^l,w^l)$. We can say that firm l is cost efficient relative to firm k if and only if $w^l.x^l=c^k(y^l,w^l)$. Otherwise, $w^l.x^l\geq c^k(y^l,w^l)$ and firm l is cost inefficient relative to firm k.

Returning to our formal definition of cost efficiency in (3.9), the index of cost efficiency of firm l relative to firm k using the production technology and cost structure of firm l is given by the ratio

$$Q(C^k,C^l,y^k,y^l,w^k,w^l) =e^l(C^k,y^k,w^k) \;/\; e^l(C^l,y^l,w^l) \tag{3.10}$$

Again, we now have two indices to compare the cost efficiency of firms

k and l - one from the point of view of firm k, and the other from the point of view of firm l. We define a symmetric index as the geometric mean of the two unsymmetric indices, thus

$$\log Q(C^k,C^l,y^k,y^l,w^k,w^l) \tag{3.11}$$
$$= \{\log[e^k(C^k,y^k,w^k)/e^k(C^l,y^l,w^l)]+\log[e^l(C^k,y^k,w^k)/e^l(C^l,y^l,w^l)]\}/2$$

As with $P(x^k,x^l,y^k,y^l)$ as a measure of technical efficiency, it will be seen that $Q(C^k,C^l,y^k,y^l,w^k,w^l)$ as a measure of cost efficiency has additional advantages when applied to a particular general class of cost function.

It is also of interest to again digress and note that an alternative measure of cost efficiency can be obtained by seeing how much each output of firm l must be multiplied by so as the firm lies on the cost function of firm k. This may be known as "output-based cost efficiency" and under constant returns to scale will equate to the "input-based cost efficiency" discussed above.

Returning to "input-based efficiency", to make further progress we may assume $\log e^s(c,y,w)$ to be a transcendental logarithmic function with the further assumption that the coefficients on the second order terms are the same for all firms (ie. independent of s). If firms k and l are producing on their cost frontiers so that $e^k(C^k,y^k,w^k)=e^l(C^l,y^l,w^l)=1$ (ie. both are efficient relative to their own production technology and cost structure), then the cost efficiency index between the two firms can be shown to be

$$\log Q(C^k,C^l,y^k,y^l,w^k,w^l) \tag{3.12}$$
$$= \frac{1}{2}\sum_{m=1}^{M}\left(T^k_m/\varepsilon^k+T^l_m/\varepsilon^l\right)\log\left(y^l_m/y^k_m\right) - \log\left(C^l/C^k\right) + \frac{1}{2}\sum_{n=1}^{N}\left(S^k_n+S^l_n\right)\log\left(w^l_n/w^k_n\right)$$

This is almost a Tornqvist cost efficiency index, except for an adjustment for local returns to scale to the revenue share of each output, reducing exactly to the Tornqvist cost efficiency index in the case of $\varepsilon^k=\varepsilon^l=1$ (constant returns to scale in both firms).

Barrow and Wagstaff (1989) contains a survey of attempts to measure cost efficiency in government health care provision and Denny and De Fontenay (1988) in telecommunications using Tornqvist cost efficiency indices.

3. Efficiency Measurement in Dynamic Cost Structure Models

In this section, we introduce time into the notation by adding a new subscript, t, to each of the elements of the input and output bundles (denoting them as x_t and y_t respectively) and each of the elements of the price vectors (denoting them as w_t and p_t respectively).

We assume that adjustments in factor usage add to the firm's cost by such things as recruitment costs, redundancy payments, external training, equipment installation, equipment commissioning or de-commissioning, and equipment sale costs. All of these adjustment costs are in addition to factor payments. This suggests that the structure of a firm's total costs are given by

$$A_t = \gamma_t C_t \tag{3.13}$$

$$\text{where } C_t = w_t.x_t = \sum_{n=1}^{N} w_{nt} x_{nt} \tag{3.14}$$

$$\text{and} \quad \gamma_t = \gamma(x_t, x_{t-1}) \tag{3.15}$$

We can be more specific about the form we expect the function $\gamma(x_t, x_{t-1})$ to take. We would expect the following conditions to be satisfied.

C1: No adjustment costs are incurred when factor inputs remain unchanged[8] ie.

$$\log \gamma(x_t,x_t) = 0 \tag{3.16}$$

C2: Adjustment costs are minimised when factor inputs remain unchanged ie.

$$\left.\frac{\partial \log\gamma}{\partial \log x_t}\right|_{x_t=x_{t-1}} = \left.\frac{\partial \log\gamma}{\partial \log x_{t-1}}\right|_{x_{t-1}=x_t} = 0$$

C3: Adjustment costs are "U" shaped in $\log x_t$ and $\log y_t$ ie.

$$\left.\frac{\partial^2 \log\gamma}{\partial \log x_t^2}\right|_{x_t=x_{t-1}}, \left.\frac{\partial^2 \log\gamma}{\partial \log x_{t-1}^2}\right|_{x_{t-1}=x_t} > 0$$

We need to study this model further for the case where the firm is technically efficient and cost minimising. This will provide a reference against which we can measure cost efficiency

We begin by introducing the firm's cost function. We assume the firm seeks to adopt an input bundle, x_t, that will minimise A_t. Therefore, the minimum cost function for firm s can be written as

$$a^s(y_t,w_t,x_{t-1}) = \min \{ A_t : x_t \in X^s(y_t) \} \tag{3.17}$$

---

[8] Since the stock of labour and equipment will naturally deplete each year, requiring new labour and equipment for the firm to maintain its level of output, it could be argued that adjustment costs should only be zero when $x_{nt}=\delta_n x_{nt-1}$, where δ_n is determined by the depreciation rate of equipment, or the rate of natural wastage of labour. An analogous feature applies to buildings as branches are opened and closed in response to migrating populations (e.g. north to south). This complication makes no difference to the final specification of the model.

where $X^s(y_t)$ is the input requirement set for firm s at time t.[9]

$a^s(y_t,w_t,x_{t-1})$ will be homogeneous of degree one in w_t, continuous, non-decreasing and concave in w_t, and non-decreasing in y_t. We assume $a^s(y_t,w_t,x_{t-1})$ is differentiable. Partial derivatives of this cost function will be used later and so are derived in Appendix A.

Note that $a^s(y_t,w_t,x_{t-1})$ encompasses both factor payments and adjustment costs, and represents (in a reduced form of only exogenous and pre-determined variables) what an efficient firm may achieve by minimising the product of these two components to cost.

It will also be noted that in (3.17) we are simply minimising the current period cost without taking into account expected discounted future costs. Therefore, we implicitly assume myopic behaviour by the firm.

It is helpful to graphically portray the cost structure we have just introduced. We do this for the case of two factor inputs (x_{1t} and x_{2t}). Suppose, for the purposes of the illustration that output is fixed at some specified level. In Figure I we show the resulting input requirement set of the firm being bounded by the iso-quant, FF. In the

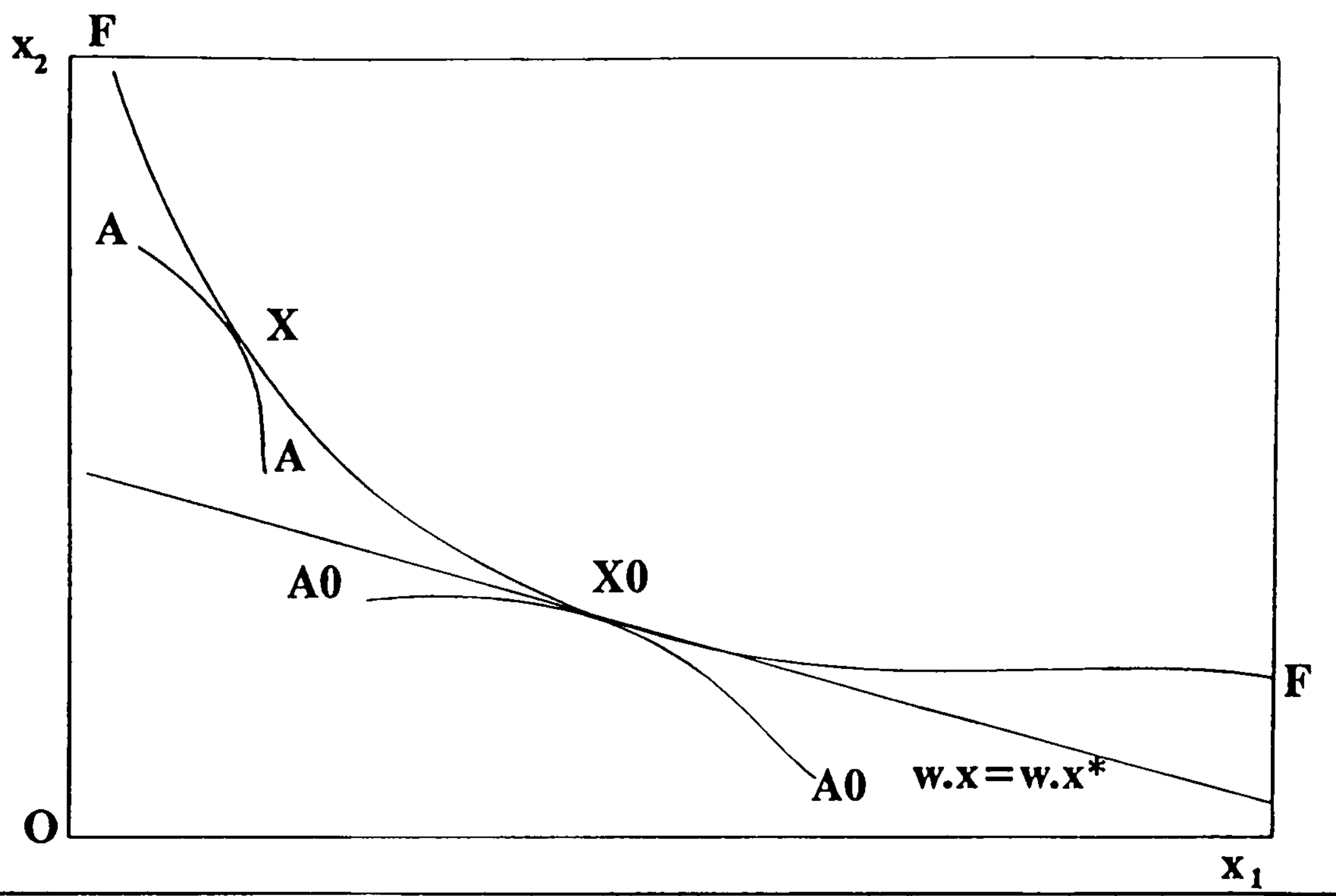
---

9 An alternative structural model could specify that adjustments to a firm's factor inputs reduce the productivity of those inputs. This specification would be appropriate if, for example, new labour needed to spend time in training, or new machines were be used at less than full productivity as they were being tested and employees trained to operate them. (However, it would not be so appropriate for some costs associated with down sizing of the firm, e.g. redundancy payments). In this case we would write the cost minimisation problem as being to minimise $w_t x_t$ subject to the constraint $f(x_t/\gamma_t,y_t)=0$ where γ_t has the same properties as before. In general, this will lead to a different cost function, $a^s(y_t,w_t,x_{t-1})$, but one that retains the same set of arguments, and can be used in a similar manner in the remaining analysis of this chapter.

absence of any adjustment costs the firm would produce at the point where the line whose gradient is given by the relative factor input prices ($-w_{1t}/w_{2t}$), the *iso-cost line*, is tangential to FF (at X_0). The effect of adjustment costs is to bend the iso-cost lines into curves, always lying below the relevant iso-cost line (in which there are no adjustments costs) since adjustment costs are always positive. Since the adjustment cost function is assumed to be concave, so the iso-cost curve will be convex. Figure I shows the case when the last period's factor inputs happen to co-incide with this period's optimal values (at X_0), giving an iso-cost curve of A_0A_0. This will be one of a series of non-intersecting curves, parameterised on the same particular level of the previous period's factor inputs (x_{t-1}).

Also shown is another (usual) case, based on a different set of previous period factor inputs, where last and this period's factor inputs do not co-incide. Here we have a different iso-cost curve, AA, with optimal factor inputs given at X. Note that since AA and A_0A_0 are conditional on different values of the previous period factor inputs, their anchorage points are different. For this reason, they can (and generally will) intersect. Clearly, there will be a set of iso-cost curves corresponding to every level or previous period factor inputs, parameterised on x_{t-1}.

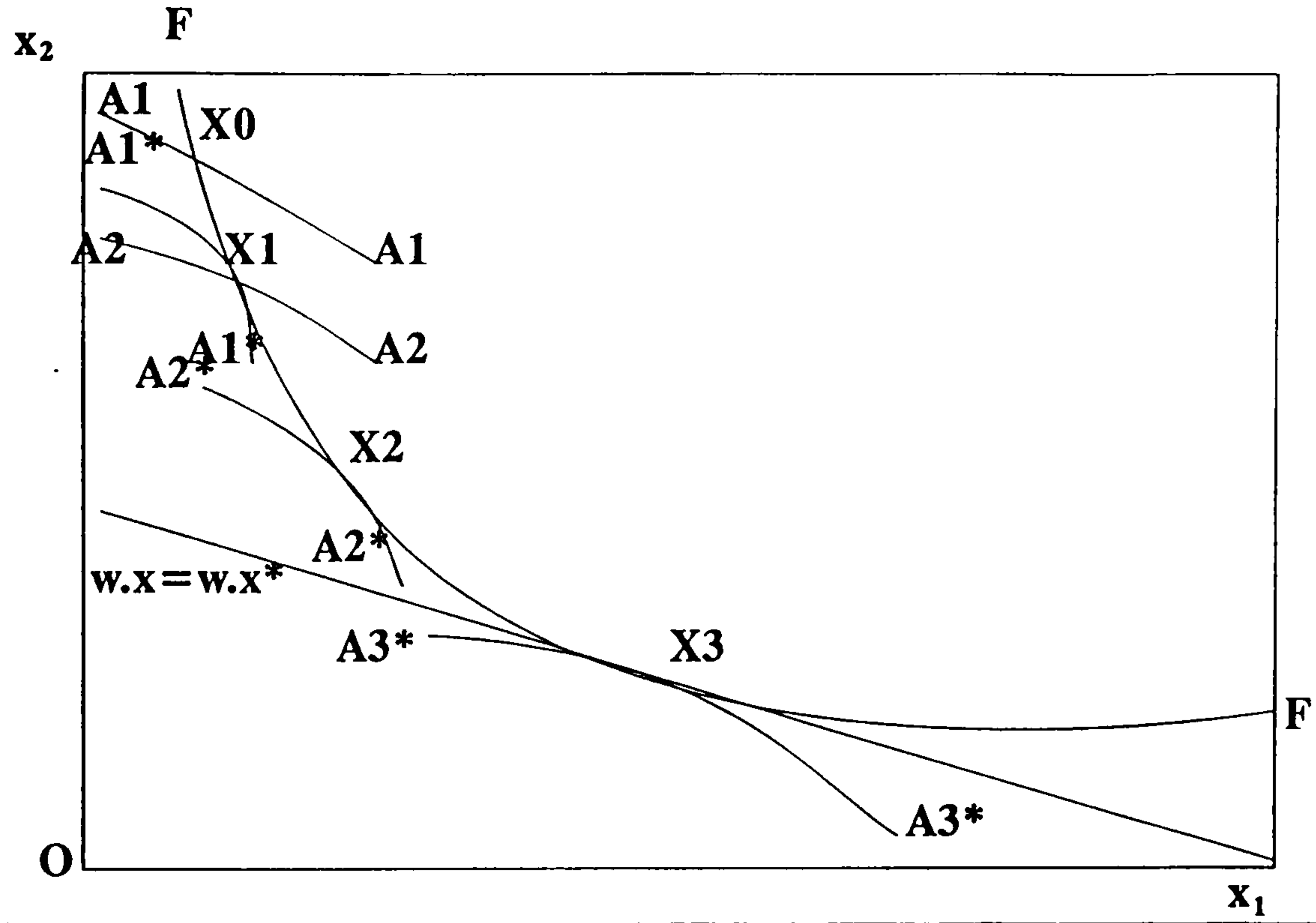
Figure I
Optimal Factor Inputs



In Figure II we assume factor prices and outputs remain unchanged between periods, with an input requirement set bounded by FF and, in the absence of adjustment costs, iso-cost lines with gradients of $-(w_{1t}/w_2)_t$. We now illustrate how factor inputs will converge to equilibrium levels equal to those in the static case of no adjustment costs. Suppose that in period *0* factor inputs are at point X_0. However, in period *1* this places the firm at a non-optimal point on an iso-cost curve A_1A_1 - non-optimal since a lower iso-cost curve, $A_1^*A_1^*$, will still allow the firm to produce within the input requirement set. The optimal input factors for period *1* are, therefore, given by the point X_1. In period *2* the iso-cost curves once again shift (since the previous period starting point from which adjustment costs are incurred has shifted) and the firm's position on the iso-cost curve A_2A_2 is no longer optimal. This prompts a change in input factors to point X_2 on the iso-cost curve $A_2^*A_2^*$. This sequence will continue until eventually factor inputs converge to point X_3 where A_3A_3 and $A_3^*A_3^*$ coincide. At this point there are no adjustment costs and the firm is in equilibrium for the given level of factor prices and output.

Figure II

Optimal Factor Inputs: Convergence to Equilibrium



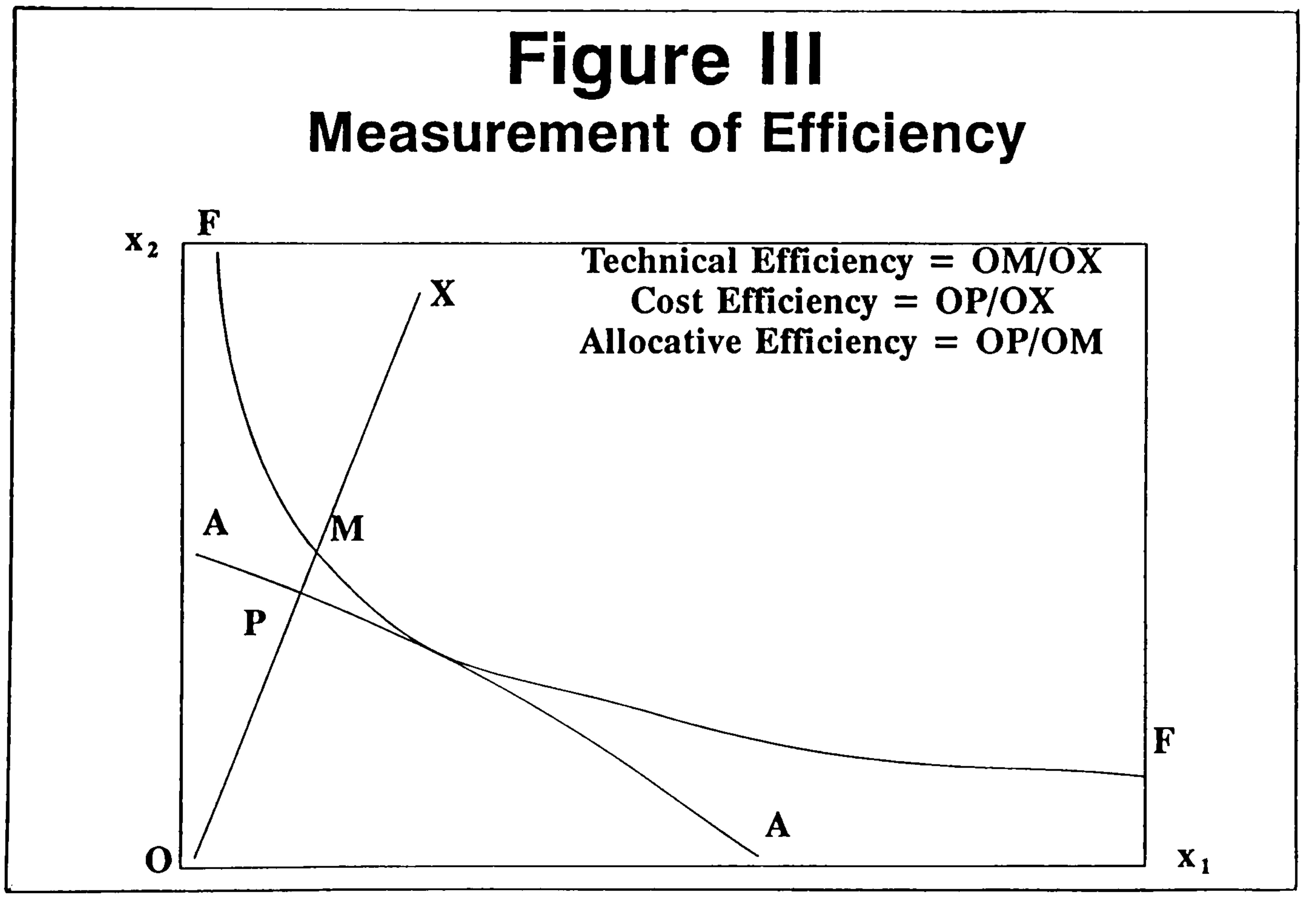
The strategy we now wish to follow is to calculate the elements of a new cost efficiency index based on the cost function $a^s(y_t,w_t,x_{t-1})$. The distance function (defined in equation (3.8)) needs to be modified to

$$e^s(A_t,y_t,w_t,x_{t-1}) = \max_\sigma \{\sigma{:}a^s(y_t,w_t,x_{t-1})\leq A_t/\sigma\} \tag{3.18}$$

Equation (3.18) requires some explanation. The distance function is equal to the scalar by which the total cost of firm s (including adjustment costs) must be divided in order that this observed cost falls on the cost frontier of firm s.

$e^s(A_t,y_t,w_t,x_{t-1})$ will be homogeneous of degree one in w_t, non-increasing and convex in w_t, and non-increasing in y_t. $e^s(A_t,y_t,w_t,x_{t-1})$ will also be differentiable.

In the case of two factor inputs (x_1 and x_2) Figure III illustrates how this distance function can be used to derive a cost efficiency measure relative to the production technology of the input requirement set bounded by FF and the iso-cost curve AA. As in the static case of no adjustment costs, we can identify technical and allocative efficiency, which combine in cost efficiency.



Analogously to the static model in equation (3.9), we can define the cost efficiency index between firm l relative to firm k's production technology and cost structure to be the ratio

$$Q^k_t(A^k_t,A^l_t,y^k_t,y^l_t,w^k_t,w^l_t,x^k_{t-1},x^l_{t-1}) = e^k(A^k_t,y^k_t,w^k_t,x^k_{t-1})/e^k(A^l_t,y^l_t,w^l_t,x^l_{t-1}) \quad (3.19)$$

which is a real number. Note that, in general, we may wish to compare firm l in period t to firm k in some other period. This would introduce a different time subscript below the cost, outputs and factor inputs and prices of firm k. However, in the interests of simplicity of notation, we assume that both firms are compared in the same period, t.

The definition of cost efficiency in equation (3.19) requires some explanation. The index compares the performance of firm l to firm k on the basis that firm l is required to produce the same output as firm k faced by the same factor prices and production technology, *and* starting from the same previous period position of inputs as firm k. We are *not* requiring firm l to incur adjustment costs from changing from its own previous period position, but we do require it to incur necessary adjustment costs on the same basis as firm k incurs when moving from its previous to current position.

Analogously to the static model in equation (3.11), we can define the cost efficiency index between firm k and l (averaged to be independent of which of the two firms is selected as the base) to be

$$\begin{aligned} \log Q_t(A^k_t,A^l_t,y^k_t,y^l_t,w^k_t,w^l_t,x^k_{t-1},x^l_{t-1}) &= \{ \log [e^k(A^k_t,y^k_t,w^k_t,x^k_{t-1})/e^k(A^l_t,y^l_t,w^l_t,x^l_{t-1})] \\ &\quad + \log [e^l(A^k_t,y^k_t,w^k_t,x^k_{t-1})/e^l(A^l_t,y^l_t,w^l_t,x^l_{t-1})] \} / 2 \end{aligned} \quad (3.20)$$

If we assume $e^s(A_t,y_t,w_t,x_{t-1})$ is a transcendental logarithmic function with second order terms held fixed across firms, then we may use

Diewert's Quadratic Identity[10] (see Diewert (1976)) to show that

$$\log Q_t(A_t^k, A_t^l, y_t^k, y_t^l, w_t^k, w_t^l, x_{t-1}^k, x_{t-1}^l)$$

$$= \quad -\frac{1}{2}\left(\left.\frac{\partial \log e^k}{\partial \log y_t}\right|_{y_t=y_t^k} + \left.\frac{\partial \log e^l}{\partial \log y_t}\right|_{y_t=y_t^l}\right)(\log y_t^l - \log y_t^k)$$

$$-\frac{1}{2}\left(\left.\frac{\partial \log e^k}{\partial \log A_t}\right|_{A_t=A_t^k} + \left.\frac{\partial \log e^l}{\partial \log A_t}\right|_{A_t=A_t^l}\right)(\log A_t^l - \log A_t^k)$$

$$-\frac{1}{2}\left(\left.\frac{\partial \log e^k}{\partial \log w_t}\right|_{w_t=w_t^k} + \left.\frac{\partial \log e^l}{\partial \log w_t}\right|_{w_t=w_t^l}\right)(\log w_t^l - \log w_t^k)$$

$$-\frac{1}{2}\left(\left.\frac{\partial \log e^k}{\partial \log x_{t-1}}\right|_{x_{t-1}=x_{t-1}^k} + \left.\frac{\partial \log e^l}{\partial \log x_{t-1}}\right|_{x_{t-1}=x_{t-1}^l}\right)(\log x_{t-1}^l - \log x_{t-1}^k) \qquad (3.21)$$

Re-writing from vector notation to summations

---

10 Diewert's Quadratic Identity states that

$$\log q(v_1) - \log q(v_2) = \frac{1}{2}\left(\left.\frac{\partial \log q}{\partial \log v}\right|_{v=v_1} + \left.\frac{\partial \log q}{\partial \log v}\right|_{v=v_2}\right).(\log v_1 - \log v_2)$$

where $\log q(v)$ is a transcendental logarithmic function (in the vector v) that evaluates to real valued number. From this it can be shown that if $\log q^1(v)$ and $\log q^2(v)$ are two transcendental logarithmic functions with the same coefficients on second order terms, then

$$\log\left(q^1(v_1)/q^1(v_2)\right) - \log\left(q^2(v_1)/q^2(v_2)\right)$$
$$= \frac{1}{2}\left(\left.\frac{\partial \log q^1}{\partial \log v}\right|_{v=v_1} + \left.\frac{\partial \log q^2}{\partial \log v}\right|_{v=v_2}\right).(\log v_1 - \log v_2)$$

$$\log Q_t(A^k_t, A^l_t, y^k_t, y^l_t, w^k_t, w^l_t, x^k_{t-1}, x^l_{t-1})$$

$$= \frac{1}{2}\sum_{m=1}^{M}\left(\left.\frac{\partial \log e^k}{\partial \log y_{mt}}\right|_{y_{mt}=y^k_{mt}} + \left.\frac{\partial \log e^l}{\partial \log y_{mt}}\right|_{y_{mt}=y^l_{mt}}\right)(\log y^l_{mt} - \log y^k_{mt})$$

$$- \frac{1}{2}\left(\left.\frac{\partial \log e^k}{\partial \log A_t}\right|_{A_t=A^k_t} + \left.\frac{\partial \log e^l}{\partial \log A_t}\right|_{A_t=A^l_t}\right)(\log A^l_t - \log A^k_t)$$

$$+ \frac{1}{2}\sum_{n=1}^{N}\left(\left.\frac{\partial \log e^k}{\partial \log w_{nt}}\right|_{w_{nt}=w^k_{nt}} + \left.\frac{\partial \log e^l}{\partial \log w_{nt}}\right|_{w_{nt}=w^l_{nt}}\right)(\log w^l_{nt} - \log w^k_{nt})$$

$$+ \frac{1}{2}\sum_{n=1}^{N}\left(\left.\frac{\partial \log e^k}{\partial \log x_{nt-1}}\right|_{x_{nt-1}=x^k_{nt-1}} + \left.\frac{\partial \log e^l}{\partial \log x_{nt-1}}\right|_{x_{nt-1}=x^l_{nt-1}}\right) \tag{3.22}$$

$$(\log x^l_{nt-1} - \log x^k_{nt-1})$$

When each firm is producing on its own cost function (ie. is cost efficient relative to its own production technology and cost structure so that $e^k(A^k_t, y^k_t, w^k_t, x^k_{t-1}) = e^l(A^l_t, y^l_t, w^l_t, x^l_{t-1}) = 1$) the partial derivatives of the distance function (see Appendix B) can be shown to be

$$\frac{\partial \log e^s}{\partial \log A_t} = 1 \tag{3.23}$$

$$\frac{\partial \log e^s}{\partial \log y_{mt}} = -T_m/\varepsilon^s_t \quad ; \quad m = 1,2,\ldots,M \tag{3.24}$$

$$\frac{\partial \log e^s}{\partial \log w_{nt}} = -S_{nt} - \frac{\partial \log \gamma_t}{\partial \log w_t} \quad ; \quad n=1,2,\ldots,N \tag{3.25}$$

$$\frac{\partial \log e^s}{\partial \log x_{nt-1}} = -\frac{\partial \log \gamma_t}{\partial \log x_{nt-1}} \quad ; \quad n=1,2,\ldots,N \tag{3.26}$$

Therefore, we can write equation (3.22) as

$$\log Q_t(A^k_t, A^l_t, y^k_t, y^l_t, w^k_t, w^l_t, x^k_{t-1}, x^l_{t-1})$$

$$= \frac{1}{2}\sum_{m=1}^{M}\left(T^k_{mt}/\varepsilon^k_t + T^l_{mt}/\varepsilon^l_t\right)(\log y^l_{mt} - \log y^k_{mt}) - \log A^l_t + \log A^k_t$$

$$+ \frac{1}{2}\sum_{n=1}^{N}\left(S^k_{nt} + S^l_n + \sum_{g=1}^{N}\eta_{gnt}\left.\frac{\partial\log\gamma}{\partial\log x_{gt}}\right|_{x_{gt}=x^k_{gt}} + \sum_{g=1}^{N}\eta_{gnt}\left.\frac{\partial\log\gamma}{\partial\log x_{gt}}\right|_{x_{gt}=x^l_{gt}}\right)$$

$$(\log w^l_{nt} - \log w^k_{nt})$$

$$+ \frac{1}{2}\sum_{n=1}^{N}\left(\left.\frac{\partial\log\gamma}{\partial\log x_{nt-1}}\right|_{x_{nt-1}=x^k_{nt-1}} + \left.\frac{\partial\log\gamma}{\partial\log x_{nt-1}}\right|_{x_{nt-1}=x^l_{nt-1}}\right)$$

$$(\log x^l_{nt-1} - \log x^k_{nt-1}) \tag{3.27}$$

where

$$T^s_m = p^s_{mt} y^s_{mt} \Big/ p^s_t . y^s_t \tag{3.28}$$

and where η_{gnt} is the cross-price elasticity for firm s at time t of the factor demand for input g with respect to the price of input n.

In the case of no adjustment costs (when $\gamma(x_t, x_{t-1}) \equiv 1$) equation (3.27) reduces to a familiar Tornqvist cost efficiency index with a modification for non-constant returns to scale. However, in the case where there are adjustment costs (when $\gamma(x_t, x_{t-1}) \neq 1$) a number of differences need to be mentioned. The costs we are comparing, A^l_t and A^k_t, now include both factor payments and adjustment costs. As in the static index, allowances are made for the implications of differences in output and factor prices between the two firms. The weights that are used to allow for different levels of output between the two firms will now depend on both the underlying degree of cost returns to scale and the optimal adjustment costs associated with any differences in output. Therefore, in allowing for an increase in output, we allow for its effect on both optimal factor payments and optimal adjustment costs. Furthermore, the weight given to factor prices in the index is modified to include the effect that a change in factor prices will have on optimal adjustment costs (via its effect on the optimal factor mix, if the price elasticities of the factor demands are non-zero). Last, the final term in the index allows for the fact that optimal adjustment costs will depend on the last period's factor input levels.

Before proceeding to the empirical work, one possible practical problem that could arise with the use of this model should be considered carefully. To see this, first note that adjustment costs have the effect of altering the unit cost of each factor input. If a factor demand changes from the previous period (either up or down), the effective unit cost of that factor increases because of adjustment costs. Similarly, under monopsony pricing with a downward sloping factor demand curve, the unit price of a factor input is related to the level of demand for that factor, such that if demand increases (decreases) the unit cost of that factor increases (decreases) so that $w_{nt}=w_{nt}(x_{nt})$ with $\partial w_{nt}/\partial x_{nt}>0$.

We have assumed that firms are price takers in their factor input markets. Suppose on the other hand that this assumption is not valid, and, in particular, that the price of a factor to a firm rises, as the firm's demand for that factor rises in any one period as would be the case under monopsony pricing. In this case either adjustment costs or a rising factor demand curve would imply rising unit costs for the factor as the firm's demand increases. Therefore, the implication for the firm's cost minimising behaviour of either an adjustment cost model or a rising factor demand curve model would be broadly the same - both predict that the unit cost of a factor input rises as demand increases. In a situation where, over the data sample, one of the factor demands <u>never</u> <u>falls</u> from the previous period, this fact could be problematic for the empirical identification of adjustment costs versus the gradient of an upward sloping factor demand curve.

Fortunately, however, there is no such problem provided all factor demands fall in at least one period during the data sample. In these periods the adjustment cost model and the upward sloping demand curve model (monopsony pricing) imply opposite movements in cost per unit of the factor input (up and down respectively). Therefore, provided that the empirical dataset contains periods for each factor input in which demand has fallen we are able to empirically distinguish between adjustment costs and the effects of an upward sloping factor demand curve.

In our case there are a number of periods when factor inputs have fallen for individual firms, and so we are able to empirically distinguish between the effects of an upward sloping demand curve (monopsony pricing) and our adjustment cost model.

4. Empirical Analysis: UK Banking Sector

a. *Measurement of Efficiency*

Unlike the static Tornqvist cost efficiency index, the modified index proposed here (in equation (3.27)) depends on the unknown adjustment cost function, $\gamma(x_t,x_{t-1})$, and its derivatives. We can, however, proceed by re-writing equation (3.27) in a form in which the unknown components of the index are grouped together in such a way as to make it susceptible to econometric estimation, with the dynamic cost efficiency index that we wish to measure forming the residual of the estimated equation. To do this we re-write equation (3.27) as

$$\log(C^l_t/C^k_t) - \frac{1}{2}\sum_{n=1}^{N}\left(S^k_{nt} + S^l_{nt}\right)\log(w^l_{nt}/w^k_{nt})$$

$$= \frac{1}{2}\sum_{m=1}^{M}\left(T^{+k}_{mt}/\varepsilon^k_t + T^{+l}_{mt}/\varepsilon^l_t\right)\left(\log y^l_{mt} - \log y^k_{mt}\right) - \log\gamma^l_t + \log\gamma^k_t$$

$$+ \frac{1}{2}\sum_{n=1}^{N}\left(\sum_{g=1}^{N}\eta^k_{gnt}\frac{\partial\log\gamma}{\partial\log x_{gt}}\bigg|_{x_{gt}=x^k_{gt}} + \sum_{g=1}^{N}\eta^l_{gnt}\frac{\partial\log\gamma}{\partial\log x_{gt}}\bigg|_{x_{gt}=x^l_{gt}}\right)(\log w^l_{nt} - \log w^k_{nt})$$

$$+ \frac{1}{2}\sum_{n=1}^{N}\left(\frac{\partial\log\gamma}{\partial\log x_{nt-1}}\bigg|_{x_{nt-1}=x^k_{nt-1}} + \frac{\partial\log\gamma}{\partial\log x_{nt-1}}\bigg|_{x_{nt-1}=x^l_{nt-1}}\right)$$

$$(\log x^l_{nt-1} - \log x^k_{nt-1})$$

$$- \log Q_{lt} \qquad (3.29)$$

where the subscript k now refers to the reference point and Q_{lt} is our cost efficiency index for firm l plus model error. Essentially, the left hand side of this equation gives total cost adjusted for input prices. The right hand side variables are output (adjusted for economies of scale), adjustment costs and various other second order terms that allow for the fact that the firm will attempt to arrange its production in such a way as to minimise adjustment costs (along with factor payments).

Equation (3.29) can be used to make efficiency comparisons between two firms (or between time periods) - l and k. However, an overall view of how a firm has performed relative to a group of firms requires the selection of a common reference point (as "firm" k). The selection of a reference point will have a bearing on the relative efficiency estimates. For the purposes of empirical analysis we take as a reference point the geometric mean of the output and each input over the sample period. That is

$$C^k_t = \bar{C} \qquad (3.30)$$

$$w^k_{nt} = \bar{w}_n \quad ; \quad n=1,2,\ldots,N \qquad (3.31)$$

$$y^{k}_{mt} = \bar{y}_{m} \quad ; \quad m=1,2,...,M \tag{3.32}$$

$$x^{k}_{nt} = \bar{x}_{n} \quad ; \quad n=1,2,...,N \tag{3.33}$$

$$S^{k}_{nt} = \bar{S}_{n} \tag{3.34}$$

where bars above variables indicate sample means.

These values are substituted for x^{k}_{t-1}, y^{k}_{t}, w^{k}_{t}, C^{k}_{t} and S^{k}_{t} in equation (3.29). We then redefine the variables as proportions of the reference as follows

$$\tilde{C}^{1}_{t} = C^{1}_{t}/\bar{C} \tag{3.35}$$

$$\tilde{w}^{1}_{nt} = w^{1}_{nt}/\bar{w}_{n} \quad ; \quad n=1,2,...,N \tag{3.36}$$

$$\tilde{y}^{1}_{mt} = y^{1}_{mt}/\bar{y}_{m} \quad ; \quad m=1,2,...,M \tag{3.37}$$

$$\tilde{x}^{1}_{nt} = x^{1}_{nt}/\bar{x}_{n} \quad ; \quad n=1,2,...,N \tag{3.38}$$

We tested the robustness of the results to changes in the reference by selecting an alternative reference point, incrementing inputs by 10%.

b. *Empirical Specification*

For empirical work we need to make an assumption about the form of the right hand side of equation (3.29), consistent with an adjustment cost function displaying non-zero second order derivatives (in logarithms) - and more precisely conditions C1, C2 and C3. We start by supposing that the adjustment cost function is quadratic in logarithms, i.e.

$$\log\gamma_{t} = \sum_{n=1}^{N} \pi_{n} (\log x_{nt} - \log x_{nt-1})^{2} \tag{3.39}$$

and the endogenous factor demands, $\log x_{nt}$ ($n=1,2,...N$), can be written as linear combinations of the exogenous and pre-determined variables, $\log y_{mt}$ $(m=1,2,...,M)$, $\log w_{nt}$ and $\log x_{nt-1}$ $(n=1,2,...,N)$. We also suppose that the inverse of the degree of returns to scale, $1/\varepsilon_t$, can be written as a linear combination of $\log y_{mt}$ $(m=1,2,...,M)$. Finally, the revenue weights, T_{mt} $(m=1,2,...,M)$, and the price elasticities, η_{nt} $(n=1,2,...,N)$, are taken to be constants over time. As we shall see, these assumptions were tested during the econometric analysis.

Appendix C shows that these assumptions allow us to re-parameterise the reduced form of equation (3.29) as a transcendental logarithmic function in the variables $\log x_{t-1}$, $\log y_t$, and $\log w_t$. In any case, such a function is known to provide a second order approximation to any arbitrary function around a specific point (such as the sample mean in the observed data set). Therefore, we can write the reduced form of equation (3.29) as

$$\begin{aligned}
\log \tilde{C}^l_t - \frac{1}{2}\sum_{n=1}^{N}\left(S^k_{nt} + \bar{S}_n \right)\log \tilde{w}^l_{nt}
&= \alpha + \sum_{m=1}^{M}\beta_m \log \tilde{y}_{mt} + \sum_{n=1}^{N}\chi_n \log \tilde{x}_{nt-1} + \sum_{n=1}^{N}\phi_n \log \tilde{w}_{nt} \\
&\quad + \sum_{m=1}^{M}\sum_{h=m}^{M}\beta_{hm}\log \tilde{y}_{ht}\log \tilde{y}_{mt} + \sum_{n=1}^{N}\sum_{g=n}^{N}\chi_{gn}\log \tilde{x}_{nt-1}\log \tilde{x}_{gt-1} \\
&\quad + \sum_{n=1}^{N}\sum_{g=n}^{N}\phi_{gn}\log \tilde{w}_{nt}\log \tilde{w}_{gt} + \sum_{m=1}^{M}\sum_{n=1}^{N}\varphi_{mn}\log \tilde{y}_{mt}\log \tilde{x}_{nt-1} \\
&\quad + \sum_{m=1}^{M}\sum_{n=1}^{N}\lambda_{mn}\log \tilde{y}_{mt}\log \tilde{w}_{nt} + \sum_{n=1}^{N}\sum_{h=1}^{N}\mu_{hn}\log \tilde{x}_{ht-1}\log \tilde{w}_{nt} \\
&\quad - \log Q_{lt} \qquad (3.40)
\end{aligned}$$

We suppose that Q_{lt} can be written as a fixed effects model. That is

$$\log Q_{lt} = \tau_t d_{1l} + \upsilon_l d_{2t} + u_{lt} \qquad (3.41)$$

where d_{1l} and d_{2t} are dummy variables for firm l and year t respectively. τ_t and υ_l are, respectively, time specific and firm specific cost efficiency effects. u_{lt} is a model residual assumed to obey the usual classical regression model assumptions. This treatment of firm specific and time specific efficiency is the same as in

Bartelsman and Dhrymes (1991).

By assuming that all the coefficients are constant throughout all firms in all years, equations (3.40) and (3.41) can be estimated by least squares. This will give estimates of relative cost efficiency for each firm and each year (τ_t and υ_l).

In the absence of adjustment costs, constant returns to scale occur when

$$\sum_{m=1}^{M} \beta_m = 1 \qquad (3.42)$$

$$\sum_{m=1}^{M} \beta_{hm} = 0 \quad ; \quad h=1,...,M \qquad (3.43)$$

$$\sum_{h=1}^{M} \beta_{hm} = 0 \quad ; \quad m=1,...,M \qquad (3.44)$$

$$\sum_{m=1}^{M} \kappa_{mn} = 0 \quad ; \quad n=1,...,N \qquad (3.45)$$

$$\sum_{m=1}^{M} \lambda_{mn} = 0 \quad ; \quad n=1,...,N \qquad (3.46)$$

c. *Application to the banking sector*

The data chosen for the empirical application of the index derived in the previous section relates to the UK banking and building society sectors. In the interests of brevity, we will usually refer to both types of firms as "banks". Application to these sectors of a dynamic index incorporating adjustment costs is particularly apt because of the high rate of growth[11] and the increasing use of information technology

---

[11] Output based GDP data show that between 1978 and 1987 the banking sector grew at an average annual real rate of 7.2% compared to 2.0% for output based GDP in for the whole economy.

within the sector.

Banks can be treated as multi-product firms selling their outputs at market prices and using various factor inputs purchased at market prices in order to provide these outputs. We will assume that banks are price takers in respect of the inputs they purchase. Of course, before proceeding to apply our dynamic efficiency index to the banking sector, we need to consider the issue of what to take as the outputs, inputs and associated prices. With respect to this, previous work in applying production and cost functions to the banking industry (e.g. Clark (1988) and Drake and Weyman-Jones (1992)) has followed two different approaches.

First, the "intermediation approach", which as its name suggests, emphasises the role of banks as financial intermediaries. Banks transform funds from retail lenders (e.g. individual depositors) and wholesale lenders (e.g. funds borrowed from other financial institutions) into assets (ie. loans from the bank). They do this with the aid of physical inputs, such as labour, equipment and branch buildings. Thus the outputs are the different types of loan products of the bank, such as bank loans to firms and individuals, mortgages, leases, etc., with associated prices of interest rates the bank receives on each of these categories of loans, mortgages or leases. The inputs are labour, equipment and branches (with associated prices of, respectively, wage rates, user costs of equipment and the user cost of branch buildings) and the various categories of the banks' funds (with associated prices of interest rates the banks have to pay to obtain these funds).

In contrast, the "production approach" emphasises the role of banks in providing a range of financial services to both depositors and borrowers alike. The provision of these services require the banks to perform certain tasks and transactions for which they require labour, equipment and a branch network. Thus, the outputs are the different loans and deposits that the banks accept (with associated prices of, respectively, the interest rates the banks receive on each of these categories of loans, and the interest rates they pay to depositors).

The inputs are just labour, equipment and branch networks (with associated prices of, respectively, wage rates, user costs of equipment and the user cost of branch buildings).

There are arguments for both approaches and the choice between them reflects the definition of output and, hence, efficiency that we wish to adopt. For this thesis, we are interested in studying how banks use their physical resources (labour, equipment and branch networks) to provide loan <u>and</u> deposit services to their customers. Therefore, we adopt the "production approach".

d. *Data*

Inputs were taken to be labour, plant and machinery, and the number of branches. Plant and machinery was measured by historic book values, revalued to constant prices. A corresponding user cost was calculated, incorporating a depreciation rate based on banks' published accounting policies. It is possible that this may over-state the true asset life in a situation where technical change makes assets obsolete before the end of their technically useful lives.[12] A full description of the sources of this data can be found in the Annex at the end of this thesis.

A number of output indicators were available from bank and society annual accounts, and returns to industry organisations (the Building Societies Association and the Committee of London and Scottish Bankers). These include the value of accounts, income and interest paid. In part each of these variables provide an indication of the level of output of the bank or society to which they pertain. A full discussion of each can also be found in the Annex at the end of this thesis. All these variables are in money values. However, the volume of the banks' outputs will be related to the real value of these

---

12 This point will be discussed in more detail during the analysis of the next chapter.

variables. Therefore, each was deflated to real terms using the Retail Prices Index (RPI).

e. *Results*

Equation (3.40) was estimated imposing the restrictions for constant returns to scale ((3.42) to (3.46)). This was done since constant returns to scale are usually assumed in static efficiency indices. Furthermore, constant returns to scale have been found in a number of empirical studies of the bank and building society sectors for all but very small banks (see, for example, Clark (1988) in the U.S. and Drake and Weyman-Jones (1992) in the U.K.), and indeed are confirmed in our own econometric analysis of chapter IV.

Due to the potentially large number of terms in equation (3.40), a parsimonious approach was adopted by restricting attention to just one output indicator. The variable selected was the value of accounts held in the bank or building society (deflated by the Retail Prices Index). This variable appeared to provide the best fit to the cost data and is often used in econometric studies of the banking sector. Restricting attention to just one output reduces the possibility of serious multicollinearity common in transcendental logarithmic models. This multicollinearity can result in unstable coefficient estimates.

Those factor inputs whose coefficients were not significant (at the 5% level) were excluded, on the basis that it could be concluded that adjustment costs for these inputs were not significant. It was found, in fact, that the only inputs for which there was evidence of significant adjustment costs were the size of branch network and the level of plant and equipment employed.

One further dummy variable was included in equation (3.40) to explain additional costs occuring in years of a merger. The merger process can consume significant management, clerical and computer resources. We may expect many of the costs to be proportional to the size of the societies involved since all members (borrowers and depositors) need to be canvassed for approval. Such merger activity has been an important feature of the sectors changing structure (see Barnes (1985)) during our sample period (and, indeed, still is). The dummy variable we used took on the value of unity for an individual building society in a year in which that society took-over (or merged with) another smaller society. The size of the coefficient indicates that a take-over or merger raises the costs of the acquiring building society by 27% in the year in which the merger of take-over occurs.

The estimation technique used for equations (3.40) and (3.41) was Estimated Generalised Least Squares (EGLS). This was preferred to OLS since it was found that there was a significant difference in the variance of the residuals for observations on banks and building societies. This would result in an OLS estimator applied to the pooled data giving inefficient coefficient estimates with biased estimates for the standard errors of the coefficients. The EGLS estimator used allowed for a different residual variance between the two sets of observations.

Table I gives the results from the EGLS estimation. *Estimation 1* is the preferred model. *Estimation 3* is a re-estimation in which the reference base has been changed in order to test the sensitivity of the results to a change in the reference technology. In *Estimation 2*, all the terms relating to previous time periods have been suppressed to test the importance of the dynamic effects in the model.

Table II displays a range of diagnostic statistics. At the 5% level of significance there is no evidence of serial correlation or heteroscedasticity in the transformed residuals of *Estimation 1* (ie. the residuals after dividing by the square root of the estimated variance). Neither is there any evidence of the model's coefficients differing between banks and building societies or over time. The latter is a particularly important test in our situation, since the assumption was made that certain quantities (revenue weights, T^{+}_{mt} $(m=1,2,...,M)$, and price elasticities, η_{gnt} $(g,n=1,2,...,N)$) were indeed constant over time.

	Reference Technology: S^k_l=0.74 (sample average) S^k_k=0.15 (" ") S^k_b=0.11 (" ")				Reference Technology: S^k_l=0.64 S^k_k=0.20 S^k_b=0.16	
	Estimation 1		Estimation 2		Estimation 3	
	Coeff.	Std.Error	Coeff.	Std.Error	Coeff.	Std.Error
α	-0.1284	0.0937	0.4392	0.0497	-0.1232	0.0942
β_a	0.3955	0.0677	0.4898	0.0792	0.3962	0.0679
β_{aa}	-0.0321	0.0214	-0.0378	-0.0236	-0.0311	0.0215
χ_b	0.5047	0.0843			0.5046	0.0849
χ_{bb}	0.0490	0.0100			0.0490	0.0100
φ_{ab}	-0.0380	0.0243			-0.0381	0.0243
μ_{bb}	0.0352	0.0289			0.0352	0.0289
λ_{ab}	0.1086	0.0429	0.1147	0.0459	0.1074	0.0431
χ_k	0.0598	0.0380			0.0592	0.0380
χ_{kk}	0.0140	0.0068			0.0139	0.0068
υ_1	0.3065	0.1608	0.9189	0.1381	0.3060	0.1613
υ_2	0.2405	0.1187	0.7447	0.0954	0.2371	0.1193
υ_3	0.0623	0.1681	0.7631	0.1348	0.0668	0.1688
υ_4	0.1900	0.1302	0.7094	0.1117	0.1986	0.1307
υ_5	-0.0015	0.0669	-0.2461	0.0600	-0.0112	0.0671
υ_6	0.4616	0.2659	-1.0928	0.0727	0.4693	0.2673
υ_7	-0.4961	0.2061	-1.4218	0.2068	-0.5111	0.2074
υ_8	-0.5834	0.3511	-1.8596	0.3971	-0.6031	0.3532
υ_9	-0.4132	0.0895	-0.8590	0.0524	-0.4328	0.0897
υ_{10}	-0.5228	0.1454	-1.4107	0.1163	-0.5299	0.1463

Table I: Estimates of Equations (3.40) and (3.41)

	Reference Technology: S_l^k=0.74 (sample average) S_k^k=0.15 (" ") S_b^k=0.11 (" ")				Reference Technology: S_l^k=0.64 S_k^k=0.20 S_b^k=0.16	
	Estimation 1		Estimation 2		Estimation 3	
	Coeff.	Std.Error	Coeff.	Std.Error	Coeff.	Std.Error
τ_{77}	-0.2854	0.0692	-0.4327	0.0839	-0.2877	0.0695
τ_{78}	0.0517	0.0600	0.0057	0.0695	0.0466	0.0602
τ_{79}	0.1487	0.0561	0.1128	0.0658	0.1463	0.0563
τ_{80}	0.1540	0.0539	0.1248	0.0639	0.1525	0.0541
τ_{81}	0.1383	0.0485	0.0965	0.0563	0.1330	0.0486
τ_{82}	0.1084	0.0428	0.0674	0.0507	0.1046	0.0429
τ_{83}	0.1106	0.0383	0.0734	0.0458	0.1109	0.0384
τ_{84}	0.0735	0.0360	0.0325	0.0432	0.0754	0.0361
τ_{85}	0.0777	0.0329	0.0475	0.0392	0.0795	0.0330
τ_{86}	0.0451	0.0304	0.0285	0.0360	0.0473	0.0304
DUMR	0.2483	0.0865	0.1815	0.1248	0.2493	0.0872
Ratio of residual variances (Building Societies / Banks):						
		1.8503		3.2579		1.8955
Standard Error		0.0570		0.0639		0.0568
R squared		0.9990		0.9982		0.9990
D-W Statistic		1.7745		0.8861		1.7771
Observations		114		114		114
Deg.of Freedom		82		88		82

Table I: Estimates of Equations (3.40) and (3.41) *(continued)*

Test	Estimation 1 Distribution	Test Statistic
AR(1)	$Chi^2(1)$	1.759
ARCH(1)	$Chi^2(1)$	1.423
Structural Stability		
Banks vs. Building Societies	F(18,64)	0.943
Data (1978-1987) vs. (1978-1986)	F(9,73)	0.444
Joint Significance of Dynamic Terms		
(χ_b, χ_{bb}, φ_{ab}, μ_{bb}, χ_k and χ_{kk})	F(6,82)	4.773
Joint Significance of Bank Dummy Variables		
($v_1, v_2, \ldots, v_{10}$)	F(10,82)	46.669
Joint Significance of Time Dummy Variables		
($\tau_{77}, \tau_{78}, \ldots, \tau_{86}$)	F(10,82)	3.903

Table II: Summary of Test Statistics for Estimation 1

In can also be seen from Tables I and II that the coefficients on the dynamic terms relating to the number of branches and level of equipment in the previous year (χ_b, χ_{bb}, φ_{ab}, μ_{bb}, χ_k and χ_{kk}) are individually (see the coefficients and standard errors in Table I) and jointly significant (see the 'F' test in Table II). It can also be seen that when these terms are suppressed, equation (3.40) suffers from serious serial correlation (see the Durbin-Watson statistic on *Estimation 2* in Table I). These two facts suggest that adjustment costs are important for these factor inputs. If the variables were omitted from equation (3.40), the implied static efficiency index would not take account of these significant dynamic effects. A bank or building society undergoing adjustment costs would appear to be less efficient if the dynamic terms were not taken into account.

From Table II we see that both sets of dummy variables (for individual banks and for individual years) are significant at the 5% level. Table III and the associated Chart I display the implied efficiency indices for the UK banking sector over the period 1978 to 1987 (based on 1987=100) as measured by the dummy variables in equation (3.40) for each year. After falls in efficiency in the last years of the 1970s, cost efficiency rose at an average rate of 2.2% p.a. between 1980 and 1987. Results are also shown for the usual static cost efficiency index for comparison. Although the general movements in the two sets of indices are very similar, the static index records a lower level of cost efficiency gain of 1.8% p.a.. This means that when adjustment costs are taken into account the efficiency gains of the banks and building societies in the sample appear to be improved. This result is not altogether surprising since these sectors have coped with dramatic increases in market size over this period and technological advances that have meant changing labour/equipment ratios have imposed additional adjustment costs, particularly since 1985 when the two lines on Chart I have converged.

	Efficiency Indices	
	Dynamic Model (Estimation 1)	Static Model (Estimation 3)
1978	95.0	99.4
1979	86.2	89.3
1980	85.7	88.3
1981	87.1	90.8
1982	89.7	93.5
1983	89.5	92.9
1984	92.9	96.8
1985	92.5	95.4
1986	95.6	97.2
1987	100.0	100.0

Table III: Implied Efficiency Indices

	Efficiency Indices (Rankings shown in brackets.)	
	Dynamic Model (Estimation 1)	Static Model (Estimation 3)
Banks		
Bank 1	99.9(2)	78.2 (2)
Bank 2	73.5 (6)	31.2 (6)
Bank 3	78.5(5)	37.1 (4)
Bank 4	93.8 (3)	36.4 (5)
Bank 5	82.6 (4)	38.5 (3)
Bank 6	100.0(1)	100.0 (1)
Standard Deviation	11.4	28.5
Building Societies		
Building Society 1	91.6 (3)	64.5 (2)
Building Society 2	100.0 (1)	100.0 (1)
Building Society 3	84.4 (4)	36.8 (4)
Building Society 4	94.1 (2)	63.8 (3)
Standard Deviation	6.5	25.9

Table IV: Implied Efficiency Indices

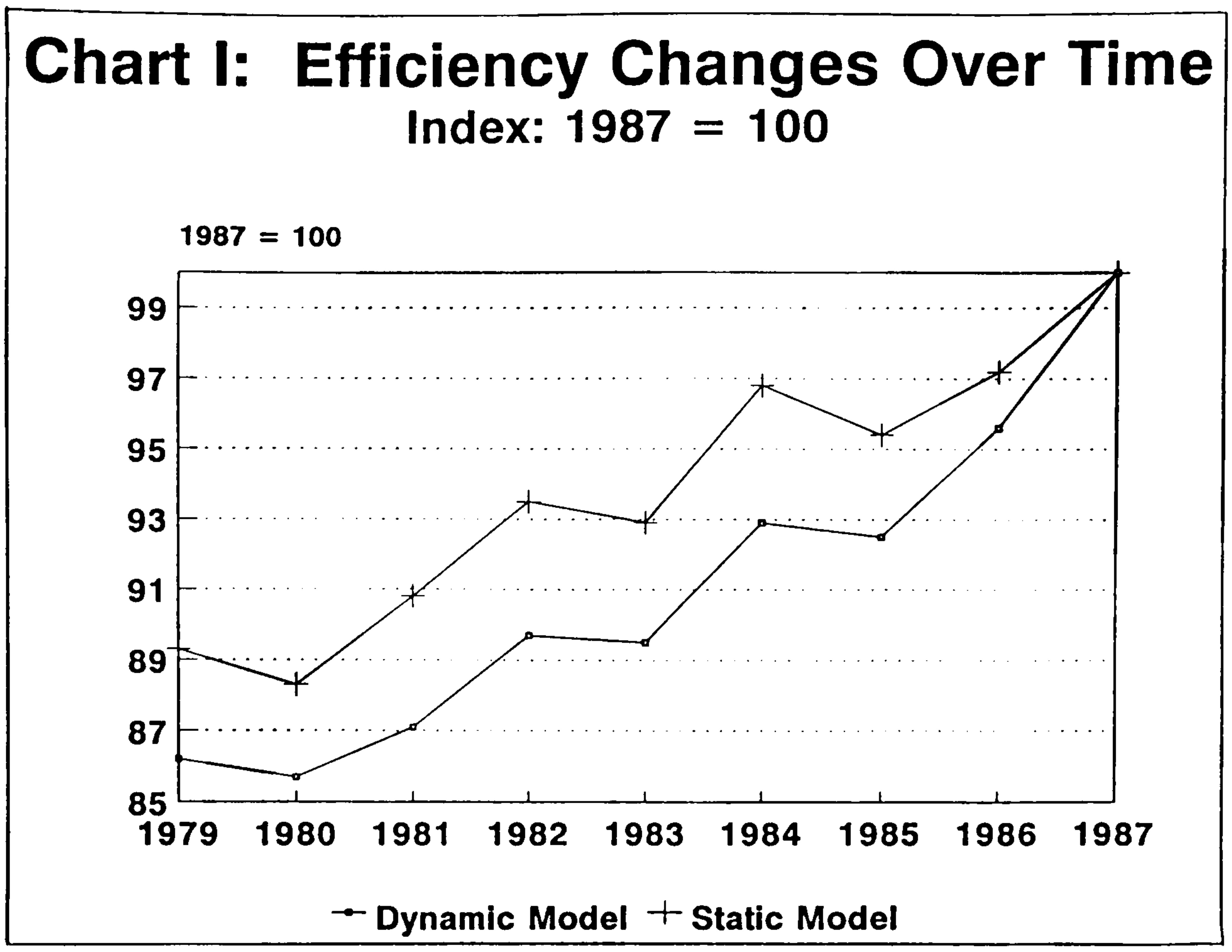


Table IV shows the results for individual banks in the sample. From inspection of the dummy variables in Table I, it is immediately clear that there are three groups of banks: Banks (1) to (6) (retail banks), Bank (7) (Standard Chartered Bank - a specialist bank) and building societies. Because of the different nature of the businesses of these groups the efficiencies measured by the dummy variables are not comparable between the groups. Therefore, a separate efficiency index has been calculated for each group, based on the most efficient bank of the group as 100.0. The first observation to make from Table IV is the large variation in efficiency performance between different building societies. This is consistent with the finding from the DEA of U.K. building societies by Drake and Weyman-Jones (1992). Table IV also shows cost efficiency indices based on a static index. Although the actual rankings of the banks and building societies do not change a great deal by the inclusion of adjustment costs in the indices, the spread of results (as shown by the standard deviation) is considerably narrower. This suggests that some inter-bank variation in cost efficiency evident from static indices may be explained by the existence of adjustment costs.

5. Conclusions

This chapter has shown the modifications needed to the usual cost efficiency indices in order to make allowance for adjustment costs affecting the optimal factor demands of a cost minimising firm. These modifications result in additional terms to the cost efficiency index which are dependent on the functional form of the adjustment costs and any fixed parameters contained within that functional form.

In an empirical example - the UK banking sector - the additional terms generated by a logarithmic adjustment cost function (depending on adjustments in the size of branch networks and the volume of equipment employed) are shown to be significant. When these terms are estimated by econometric methods, the conclusions with regard to the cost efficiency ranking of banks and building societies within those sectors do not change dramatically since all banks and building societies have grown over this period. However, the degree of variation in performance is considerably narrowed, suggesting that some of the observed differences in our cost efficiency estimates between banks and building societies results from adjustment costs. Furthermore, once account is taken of the adjustment costs suffered by the sector during this period of substantial growth and change, the estimated gain in cost efficiency of the sector as a whole since 1980 increases from an average of 1.8% p.a. to 2.2% p.a.. To the extent that these costs are an unavoidable consequence of necessary changes in scale and methods of service provision, they should not be regarded as *managerial* or allocative inefficiency.

However, there are limitations to the analysis of this chapter. First, we have not investigated the long-run equilibrium level of factor demands and costs, ie. what they would be if the firm was in a steady state incurring no adjustment costs. Knowledge of this long-term equilibrium is necessary in order to understand the importance of adjustment costs to the sector.

Second, we have assumed myopic behaviour by the firms in the sector. It may be that expectations of future demands play an important role, as firms choose current period factor demands that minimise the expected value of current and future costs.

Consideration of both these points requires the construction and explicit solution of a structural model of the firms' costs - the subject of the next chapter.

APPENDIX A: PARTIAL DERIVATIVES OF A DYNAMIC COST FUNCTION

In this appendix we derive the partial derivatives of the dynamic cost function, $a^s(y_t,w_t,x_{t-1})$. These results are analogous to Shephard's Lemma for single period static cost functions.

Let x_t^* be the cost minimising bundle that produces y_t at prices w_t when inputs in the previous period are x_{t-1} respectively, and $C_t^*=w_t.x_t^*$ is the corresponding minimum cost. Let

$$g^s(w_t^-,x_{t-1}^-) = a^s(y_t,w_t^-,x_{t-1}^-) - \gamma(x_t^*,x_{t-1}^-)\ w_t^-.x_t^* \qquad (3.A1)$$

Since, in the absence of static allocative inefficiency, $a^s(y_t,w_t^-,x_{t-1}^-)$ is the cheapest way to produce y_t, this function is always non-positive. At $(w_t^-,x_{t-1}^-)=(w_t,x_{t-1})$, $g^s(w_t,x_{t-1})=0$. Since this is a maximum value of $g^s(w_t^-,x_{t-1}^-)$, its derivatives must vanish. Differentiating with respect to w_t

$$\frac{\partial g^s}{\partial w_t} = \frac{\partial a^s}{\partial w_t} - \gamma x_t^* - \frac{\partial \gamma}{\partial w_t} w_t.x_t^* = 0 \qquad (3.A2)$$

$$\Rightarrow \quad \frac{\partial a^s}{\partial w_t} = \gamma_t x_t^* + \frac{\partial \gamma}{\partial w_t} w_t.x_t^* \qquad (3.A3)$$

Differentiating with respect to x_{t-1}

$$\frac{\partial g^s}{\partial x_{t-1}} = \frac{\partial a^s}{\partial x_{t-1}} - \frac{\partial \gamma}{\partial x_{t-1}} w_t.x_t^* = 0 \qquad (3.A4)$$

$$\Rightarrow \quad \frac{\partial a^s}{\partial x_{t-1}} = \frac{\partial \gamma}{\partial x_{t-1}} w_t.x_t^* \qquad (3.A5)$$

The final partial derivative of the cost function (a^s) that we need is that with respect to the outputs, y_t. This requires a different approach from the one adopted for the other partial derivatives since varying y_t in isolation will conflict with the transformation function constraint $f^s(x_t,y_t)=0$. We proceed as follows.

Assume that the firm chooses its output bundle, y_t, in order to maximise its revenue whilst using a given input bundle, x_t. In mathematical notation we wish to find y_t that maximises $p_t.y_t-\gamma_t w_t.x_t$ subject to the constraint $A_t=a^s(y_t,w_t,x_{t-1})$.

The first order conditions to this problem are

$$p_t - \lambda \frac{\partial a^s}{\partial y_t} = 0 \tag{3.A6}$$

$$\Rightarrow \quad \lambda \frac{\partial a^s}{\partial y_t} = p_t - C_t \frac{\partial \gamma}{\partial y_t} \tag{3.A7}$$

subject to the constraint $A_t=a^s(y_t,w_t,x_{t-1})$. λ is a Lagrange Multiplier.

Multiplying each of these equations by the corresponding element of y_t and summing gives

$$\lambda \frac{\partial f^s}{\partial y_t} .y_t = p_t.y_t - C_t \frac{\partial \gamma}{\partial y_t} .y_t \tag{3.A8}$$

Dividing (3.A7) by (3.A8) in order to remove λ we get

$$\frac{\partial a^s}{\partial y_t} \Big/ \frac{\partial a^s}{\partial y_t} .y_t = p_t \,/\, p_t.y_t \tag{3.A9}$$

In a similar manner as before, we define local cost function returns to scale for firm s as follows. Consider increasing A_t by a factor υ. Let $v^s(A_t,y_t,w_t,x_{t-1},\upsilon)$ be the factor of proportionality by which all outputs, y_t, must be increased so that the inflated input and output vectors lie on the new cost function for firm s; ie. $v^s(A_t,y_t,w_t,x_{t-1},\upsilon)$ is the solution to $\upsilon A_t=a^s(v^s y_t,w_t,x_{t-1})$. Returns to scale can be determined by considering the sensitivity of $v^s(A_t,y_t,w_t,x_{t-1},\upsilon)$ to changes in ν. Differentiating $\upsilon A_t=a^s(v^s y_t,w_t,x_{t-1})$ with respect to υ gives

$$A_t \quad = y_t.\frac{\partial a^s}{\partial y_t}\frac{\partial v^s}{\partial \upsilon} \tag{3.A10}$$

$$\frac{\partial v^s}{\partial \upsilon} \quad = A_t \Big/ y_t.\frac{\partial a^s}{\partial y^s} \tag{3.A11}$$

The degree of local cost function returns to scale is given by

$$\varepsilon_t^s = \frac{\partial v^s(A_t,y_t,w_t,x_{t-1},\upsilon)}{\partial \upsilon} \text{ evaluated at } \upsilon = 1. \tag{3.A12}$$

Therefore, local cost functions returns to scale are given by

$$\varepsilon_t^s = A_t \Big/ y_t.\frac{\partial a^s}{\partial y_t} \tag{3.A13}$$

evaluated at $\upsilon=1$ (so as to give returns to scale in the locality of A_t).

If local returns to scale are constant (increasing, decreasing), then $\varepsilon_t^s = 1$ (>1 or <1 respectively). Note that ε_t^s will depend crucially on adjustment costs faced by the firm at time *t*.

Panzar (1989) shows that, in a static model, when firms produce in a cost efficient manner, local returns to scale as defined in this way (cost function returns to scale) equate to local returns to scale defined in footnote 5 (production function returns to scale).

Using equation (3.A13) we can re-write (3.A9) as

$$\frac{\varepsilon_t^s}{A_t}\frac{\partial a^s}{\partial y_t} \quad = p_t \Big/ p_t.y_t \tag{3.A14}$$

$$\Rightarrow \frac{\partial a^{s}}{\partial y_{t}} = A_{t}\, p_{t} \,/\, [\ \varepsilon_{t}^{s}\, (p_{t}.y_{t})\ ] \tag{3.A15}$$

APPENDIX B: PARTIAL DERIVATIVES OF THE DISTANCE FUNCTION

We are interested in the partial derivatives of $e^s(A_t,y_t,w_t,x_{t-1})$ with respect to the scalar A_t and the vectors y_t, w_t, and x_{t-1}. To derive these we first use the Implicit Function Theorem to write the cost function as

$$\frac{A_t}{e^s(A_t,y_t,w_t,x_{t-1})} = a^s(y_t,w_t,x_{t-1}) \tag{3.B1}$$

Equation (3.B1) re-expresses the definition of $e^s(A_t,y_t,w_t,x_{t-1})$ as the scalar by which the firm's actual cost (A_t) exceeds that which would be achievable if the firm was efficient and cost minimising.

Differentiating (3.B1) with respect to A_t gives

$$\left(e^s(A_t,y_t,w_t,x_{t-1})-A_t\frac{\partial e^s}{\partial A_t}\right)\Big/\left(e^s(A_t,y_t,w_t,x_{t-1})\right)^2 = 0 \tag{3.B2}$$

$$\Rightarrow \quad \frac{\partial e^s}{\partial A_t} = \frac{e^s(A_t,y_t,w_t,x_{t-1})}{A_t} \tag{3.B3}$$

$$\Rightarrow \quad \frac{\partial \log e^s}{\partial \log A_t} = 1 \tag{3.B4}$$

Differentiating (3.B1) with respect to y_t gives

$$-A_t\frac{\partial e^s}{\partial y_t}\Big/\left(e^s(A_t,y_t,w_t,x_{t-1})\right)^2 = \frac{\partial a^s}{\partial y_t} \tag{3.B5}$$

$$\Rightarrow \frac{\partial e^s}{\partial y_t} = -\frac{(e^s(A_t,y_t,w_t,x_{t-1}))^2}{\varepsilon_t^s} p_t \Big/ p_t.y_t$$

(from (3.A15))

$$\Rightarrow \frac{\partial e^s}{\partial y_{mt}} = -\frac{(e^s(A_t,y_t,w_t,x_{t-1}))^2}{\varepsilon_t^s} p_{mt} \Big/ p_t.y_t \quad ; \quad m = 1,2,\ldots,M$$

$$\Rightarrow \frac{\partial \log e^s}{\partial \log y_{mt}} = \frac{e^s(A_t,y_t,w_t,x_{t-1})}{\varepsilon_t^s} p_{mt}y_{mt} \Big/ p_t.y_t \quad ; \quad m = 1,2,\ldots,M$$

$$= \frac{e^s(A_t,y_t,w_t,x_{t-1})}{\varepsilon_t^s} T_m \quad ; \quad m = 1,2,\ldots,M \tag{3.B6}$$

where

$$T_m = p_{mt}y_{mt} \,/\, p_t.y_t \tag{3.B7}$$

ie. the shares of revenue earnt by each output.

Differentiating (3.B1) with respect to w_t gives

$$-A_t \frac{\partial e^s}{\partial w_t} \Big/ (e^s(A_t,y_t,w_t,x_{t-1}))^2 = \frac{\partial a^s}{\partial w_t} = \gamma_t x_t^* + \frac{\partial \gamma}{\partial w_t} w_t.x_t^* \tag{3.B8}$$

(from(3.A3))

$$\Rightarrow \frac{\partial e^s}{\partial w_t} = -(e^s(A_t,y_t,w_t,x_{t-1}))^2 \left(\frac{x_t^*}{w_t.x_t^*} + \frac{\partial \gamma}{\partial w_t}\frac{1}{\gamma_t}\right) \tag{3.B9}$$

$$\Rightarrow \frac{\partial e^s}{\partial w_{nt}} = -(e^s(A_t,y_t,w_t,x_{t-1}))^2 \left(\frac{x_{nt}^*}{w_t.x_t^*} + \frac{\partial \gamma}{\partial w_{nt}}\frac{1}{\gamma_t}\right) \quad ; \quad n=1,2,\ldots,N$$

$$\Rightarrow \quad \frac{\partial \log e^{s}}{\partial \log w_{nt}} = - e^{s}(A_{t},y_{t},w_{t},x_{t-1}) \left(\frac{w_{nt}x^{*}_{nt}}{w_{t}.x^{*}_{t}} + \frac{\partial \log \gamma}{\partial \log w_{nt}}\right); n=1,2,\ldots,N \tag{3.B10}$$

Finally, (3.B1) differentiating with respect to x_{t-1} gives

$$- A_{t}\frac{\partial e^{s}}{\partial x_{t-1}} / (e^{s}(A_{t},y_{t},w_{t},x_{t-1}))^{2} = \frac{\partial a^{s}}{\partial x_{t-1}} = \frac{\partial \gamma}{\partial x_{t-1}} \, w_{t}.x^{*}_{t} \tag{3.B11}$$

(from(3.A5))

$$\Rightarrow \quad \frac{\partial e^{s}}{\partial x_{t-1}} = - (e^{s}(A_{t},y_{t},w_{t},x_{t-1}))^{2} \, \frac{w_{t}.x^{*}_{t}\partial \gamma}{w_{t}.x_{t}\partial x_{t-1}} \, \frac{1}{\gamma_{t}} \tag{3.B12}$$

$$\Rightarrow \quad \frac{\partial e^{s}}{\partial x_{nt-1}} = - (e^{s}(A_{t},y_{t},w_{t},x_{t-1}))^{2} \, \frac{w_{t}.x^{*}_{t}}{w_{t}.x_{t}} \, \frac{\partial \gamma}{\partial x_{nt-1}} \, \frac{1}{\gamma} \quad ; \quad n=1,2,\ldots,N$$

$$\Rightarrow \frac{\partial \log e^{s}}{\partial \log x_{nt-1}} = - e^{s}(C_{t},y_{t},w_{t},x_{t-1}) \, \frac{w_{t}.x^{*}_{t}}{w_{t}.x_{t}} \, \frac{\partial \log \gamma}{\partial \log x_{nt-1}} \quad ; \quad n=1,2,\ldots,N \tag{3.B13}$$

When the firm is producing on its own cost function so that $e^{s}(C_{t},y_{t},w_{t},x_{t-1})=1$ (ie. is cost-efficient relative to its own production technology and cost structure), then partial derivatives of the distance function ((3.B6), (3.B10) and (3.B13)) reduce to

$$\frac{\partial \log e^{s}}{\partial \log y_{mt}} = - T_{m} / \varepsilon^{s}_{t} \quad ; \quad m = 1,2,\ldots,M \tag{3.B17}$$

$$\frac{\partial \log e^{s}}{\partial \log w_{nt}} = - S_{nt} - \frac{\partial \log \gamma}{\partial \log w_{t}} \quad ; \quad n = 1,2,\ldots,N \tag{3.B18}$$

$$\frac{\partial \log e^{s}}{\partial \log x_{nt-1}} = - \frac{\partial \log \gamma}{\partial \log x_{nt-1}} \quad ; \quad n = 1,2,\ldots,N \tag{3.B19}$$

APPENDIX C: DERIVATION OF THE EMPIRICAL SPECIFICATION

Equation (3.39) tells us that

$$\log\gamma^{l}_{t} - \log\gamma^{k}_{t} = \sum_{n=1}^{N} \pi_{n} (\log x^{l}_{nt} - \log x^{l}_{nt-1})^{2} - \sum_{n=1}^{N} \pi_{n} (\log x^{k}_{nt} - \log x^{k}_{nt-1})^{2}$$

$$\begin{aligned}
&= \sum_{n=1}^{N} \pi_{n} (\log x^{l}_{nt} - \log x^{k}_{nt})^{2} + \sum_{n=1}^{N} \pi_{n} (\log x^{l}_{nt-1} - \log x^{k}_{nt-1})^{2} \\
&- 2 \sum_{n=1}^{N} \pi_{n} (\log x^{l}_{nt} - \log x^{k}_{nt})(\log x^{l}_{nt-1} - \log x^{k}_{nt-1}) \\
&+ 2 \sum_{n=1}^{N} \pi_{n} \log x^{k}_{nt}(\log x^{l}_{nt} - \log x^{k}_{nt}) + 2 \sum_{n=1}^{N} \pi_{n} \log x^{k}_{nt-1}(\log x^{l}_{nt-1} - \log x^{k}_{nt-1}) \\
&- 2 \sum_{n=1}^{N} \pi_{n} \log x^{k}_{nt-1}(\log x^{l}_{nt} - \log x^{k}_{nt}) - 2 \sum_{n=1}^{N} \pi_{n} \log x^{k}_{nt}(\log x^{l}_{nt-1} - \log x^{k}_{nt-1})
\end{aligned}$$

So using the definitions in equations (3.33) and (3.38)

$$\begin{aligned}
\log\gamma^{l}_{t} - \log\gamma^{k}_{t} &= \sum_{n=1}^{N} \pi_{n}(\log\tilde{x}^{l}_{nt})^{2} + \sum_{n=1}^{N} \pi_{n}(\log\tilde{x}^{l}_{nt-1})^{2} - 2 \sum_{n=1}^{N} \pi_{n} \log\tilde{x}^{l}_{nt} \log\tilde{x}^{l}_{nt-1} \\
&+2 \sum_{n=1}^{N} \pi_{n} \log\bar{x}_{nt} \log\tilde{x}^{l}_{nt} + 2 \sum_{n=1}^{N} \pi_{n} \log\bar{x}_{nt-1} \log\tilde{x}^{l}_{nt-1} \\
&-2 \sum_{n=1}^{N} \pi_{n} \log\bar{x}_{nt-1} \log\tilde{x}^{l}_{nt} - 2 \sum_{n=1}^{N} \pi_{n} \log\bar{x}_{nt} \log\tilde{x}^{l}_{nt-1}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{n=1}^{N} \pi_{n}(\log\tilde{x}^{l}_{nt})^{2} + \sum_{n=1}^{N} \pi_{n}(\log\tilde{x}^{l}_{nt-1})^{2} - 2 \sum_{n=1}^{N} \pi_{n} \log\tilde{x}^{l}_{nt} \log\tilde{x}^{l}_{nt-1} \\
&+2 \sum_{n=1}^{N} \pi_{n}(\log\bar{x}_{nt} - \log\bar{x}_{nt-1})\log\tilde{x}^{l}_{nt} - 2 \sum_{n=1}^{N} \pi_{n}(\log\bar{x}_{nt} - \log\bar{x}_{nt-1})\log\tilde{x}^{l}_{nt-1}
\end{aligned}$$

The term of equation (3.29) involving the partial derivatives of the adjustment cost function becomes

$$\sum_{g=1}^{N} \eta_{gnt} \frac{\partial \log\gamma}{\partial \log x_{gt}}\bigg|_{x_{gt}=x^{k}_{gt}} + \sum_{g=1}^{N} \eta_{gnt} \frac{\partial \log\gamma}{\partial \log x_{gt}}\bigg|_{x_{gt}=x^{l}_{gt}}$$

$$= 2 \sum_{g=1}^{N} \eta_{gnt} \pi_{g}(\log x^{l}_{gt} - \log x^{l}_{gt-1}) + 2 \sum_{g=1}^{N} \eta_{gnt} \pi_{g}(\log x^{k}_{gt} - \log x^{k}_{gt-1})$$

$$= 2 \sum_{g=1}^{N} \eta_{gnt} \pi_g (\log x^1_{gt} - \log x^k_{gt}) + 2 \sum_{g=1}^{N} \eta_{gnt} \pi_g (\log x^1_{gt-1} - \log x^k_{gt-1})$$

So using the definitions in equations (3.33) and (3.38)

$$\sum_{g=1}^{N} \eta_{gnt} \frac{\partial \log \gamma}{\partial \log x_{gt}} \bigg|_{x_{gt}=x^k_{gt}} + \sum_{g=1}^{N} \eta_{gnt} \frac{\partial \log \gamma}{\partial \log x_{gt}} \bigg|_{x_{gt}=x^1_{gt}}$$
$$= 2 \sum_{g=1}^{N} \eta_{gnt} \pi_g \log \tilde{x}^1_{gt} + 2 \sum_{g=1}^{N} \eta_{gnt} \pi_g \log \tilde{x}^1_{gt-1}$$

Similarly

$$\frac{\partial \log \gamma}{\partial \log x_{nt-1}} \bigg|_{x_{nt-1}=x^k_{nt-1}} + \frac{\partial \log \gamma}{\partial \log x_{nt-1}} \bigg|_{x_{nt-1}=x^1_{nt-1}} = 2\pi_n \log \tilde{x}^1_{nt} + 2\pi_n \log \tilde{x}^1_{nt-1}$$

We will specify factor demands to be linearly related to the logarithms of output, factor prices and the previous period's factor demands. That is

$$\log \tilde{x}^1_{nt} = a_n + \sum_{m=1}^{M} b_{nm} \log \tilde{y}^1_{mt} + \sum_{g=1}^{N} c_{ng} \log \tilde{w}^1_{nt} + \sum_{g=1}^{N} d_{ng} \log \tilde{x}^1_{nt-1}$$

We will also specify that the revenue weights are constant and the degree of returns to scale can be written as a linear combination of the outputs. So we write

$$T^k_{mt}/\varepsilon^k_t + T^1_{mt}/\varepsilon^1_t = e_m + \sum_{h=1}^{M} f_{mh} \log \tilde{y}^1_{ht}$$

We now only need to make the assumption that the price elasticities are constants ($\eta_{ngt}=\eta_{ng}$, $n,g=1,2,\dots,N$) to make the following re-parameterisation

$$\alpha = \sum_{n=1}^{N} \pi_n a_n^2 + 2 \sum_{n=1}^{N} \pi_n a_n \log(\bar{x}_{nt}/\bar{x}_{nt-1})$$

$$\beta_m = e_m/2 + \sum_{n=1}^{N} \pi_n a_n b_{nm} + 2 \sum_{n=1}^{N} \pi_n b_{nm} \log(\bar{x}_{nt}/\bar{x}_{nt-1})$$

$$\chi_n = \sum_{g=1}^{N} \pi_g a_g d_{gn} - 2\pi_n a_n + 2 \sum_{g=1}^{N} \pi_g d_{gn} \log(\bar{x}_{gt}/\bar{x}_{gt-1}) - 2\pi_n \log(\bar{x}_{nt}/\bar{x}_{nt-1}) - \pi_n a_n$$

$$\phi_n = \sum_{g=1}^{N} \pi_g a_g c_{gn} + 2 \sum_{g=1}^{N} \pi_g c_{gn} \log(\bar{x}_{gt}/\bar{x}_{gt-1}) + \sum_{g=1}^{N} \eta_{gn} \pi_g a_g$$

$$\beta_{hm} = f_{mh}/2 + \sum_{n=1}^{N} \pi_n b_{nm} b_{nh}$$

$$\chi_{gn} = \sum_{h=1}^{N} \pi_h d_{hg} d_{hn} + 2\Delta_{ng} \pi_n - \pi_n d_{ng}$$

$$\phi_{gn} = \sum_{h=1}^{N} \pi_h c_{hg} c_{hn} + \sum_{h=1}^{N} \eta_{hn} \pi_h c_{hg}$$

$$\varphi_{mn} = \sum_{g=1}^{N} \pi_g b_{gm} d_{gn} - 2\pi_n b_{nm} + \pi_n b_{nm}$$

$$\lambda_{mn} = \sum_{g=1}^{N} \pi_g b_{gm} c_{gn} + \sum_{g=1}^{N} \eta_{gn} \pi_g b_{gm}$$

$$\mu_{gn} = \sum_{h=1}^{N} \pi_h c_{hg} d_{hn} - 2\pi_g c_{gn} + \sum_{h=1}^{N} \eta_{hn} \pi_h d_{gh} + \eta_{gn} \pi_g + \pi_n c_{ng}$$

for $n,g=1,2,\ldots,N$ and $m,h=1,2,\ldots,M$

and where $\Delta_{gn} = 1$ when $g=n$
$= 0$ otherwise.

So that

$$
\begin{aligned}
&\log(\tilde{C}^{1}_{t}) - \frac{1}{2}\sum_{n=1}^{N}\left(S^{k}_{nt} + \bar{S}_{n}\right)\log(\tilde{w}^{1}_{nt}) \\
&= \alpha + \sum_{m=1}^{M}\beta_{m}\log\tilde{y}_{mt} + \sum_{n=1}^{N}\chi_{n}\log\tilde{x}_{nt-1} + \sum_{n=1}^{N}\phi_{n}\log\tilde{w}_{nt} \\
&\quad + \sum_{m=1}^{M}\sum_{h=m}^{M}\beta_{hm}\log\tilde{y}_{ht}\log\tilde{y}_{mt} + \sum_{n=1}^{N}\sum_{g=n}^{N}\chi_{gn}\log\tilde{x}_{nt-1}\log\tilde{x}_{gt-1} \\
&\quad + \sum_{n=1}^{N}\sum_{g=n}^{N}\phi_{gn}\log\tilde{w}_{nt}\log\tilde{w}_{gt} + \sum_{m=1}^{M}\sum_{n=1}^{N}\varphi_{mn}\log\tilde{y}_{mt}\log\tilde{x}_{nt-1} \\
&\quad + \sum_{m=1}^{M}\sum_{n=1}^{N}\lambda_{mn}\log\tilde{y}_{mt}\log\tilde{w}_{nt} + \sum_{n=1}^{N}\sum_{h=1}^{N}\mu_{hn}\log\tilde{x}_{ht-1}\log\tilde{w}_{nt} \\
&\quad - \log Q_{1t}
\end{aligned}
$$

CHAPTER IV

MEASUREMENT OF EFFICIENCY IN THE UK BANKING SECTOR BY MEANS OF A DYNAMIC MODEL OF FACTOR DEMANDS AND COST STRUCTURE

ABSTRACT

This chapter obtains a closed solution for factor demands derived from a cost structure that includes adjustment costs. This is done whilst taking account of observation difficulties in capital stock and user costs owing to the use of inappropriate depreciation rates in company accounting procedures. It is argued that this model is particularly appropriate to studying the banking sector and is used as a basis for estimating efficiency changes over time and between UK banks and building societies.

1. Introduction

In this chapter we turn to a parametric modelling approach to the measurement of efficiency.

Static models of factor demands and cost structures have been extensively studied and applied to most production industries. Numerous examples are given in Fuss and McFadden (1978) and Apostolakis (1988), the latter concentrating particularly on transcendental logarithmic cost function models. A steady stream of empirical work has continued to be produced over the last decade applying these models to a variety of industries. For just two examples, see Caves, Christensen and Tretheway (1984) and Sickles and Streitwieser (1989) dealing with the airline and natural gas industries respectively.

This chapter presents modifications to these models in two directions. First, we introduce adjustment costs into the short term factor demand equations. Early attempts at this introduced a partial adjustment mechanism into the factor demand equations without relating this to any underlying adjustment cost specification (e.g. Nissim (1982)). More recently, Hunt and Lynk (1989) used the notion of a long-run equilibrium cost and used the correspondence between error correction and co-integration mechanisms to track the adjustment to this equilibrium. These approaches do not employ a precise economic theory to support the specifications and, therefore, do not guarantee that the cost function and factor demand equations imply a reasonable production function.

There is, however, a history of rigorous theoretical derivations of factor demands in the presence of adjustment costs stretching back a number of years; the principle problem being to devise a closed analytical expression for the factor demands from a profit maximisation or cost minimisation objective. Morrison and Berndt (1981) use an approach following Treadway (1974) and Berndt, Fuss and Waverman (1977 and 1980). This leads them to introduce an adjustment cost on the quasi-fixed factor inputs by specifying a production function in which output is foregone if these inputs are varied. Assumptions are needed for the expectations of the exogenous variables. Morrison and Berndt assume that input prices are all expected to grow at the same fixed rate (thus staying constant relative to each other) and output is expected to stay constant indefinitely. They then solve a two stage minimisation problem for the present value of current and future costs, first with respect to variable inputs - the solution to which they write as a transcendental logarithmic function (which includes fixed values for the quasi-fixed inputs) - and, second, with respect to the quasi-fixed inputs themselves. The authors note that a closed solution for the case of more than one quasi-fixed input is complex. A further case study by the same authors is found in Morrison and Berndt (1991), and essentially the same model is also used by Lawrence (1990) to study short and long term export-supply and import-demand price elasticities.

Epstein (1981) notes two limitations to the Morrison and Berndt approach. First, it is only practical when there is a single quasi-fixed input. Second, the expectations assumptions are very restrictive. For this reason other authors have attempted modified approaches. Epstein (1981) proposes an approach that allows more than one quasi-fixed input by establishing a duality between the firm's technology and value functions (the maximum value of the integral of current and discounted profits). This duality can be exploited in empirical work so that dynamic factor demands can be derived from the value function in cases where explicit solutions derived directly from the technology (e.g. the production function) would be excessively complicated. However, Epstein and Yatchew (1985) make a useful contribution for the case of a quadratic production function through re-parameterising the model so that closed form analytical expressions for the factor demands can be written down. This approach is also used by Madan and Prucha (1989).

Pindyck and Rotemberg (1983) adopt an approach that essentially minimises costs in the current planning period only, and replaces the expectations of the future values of variables by their actual future values. This gives a system of equations (comprising of cost function, factor share equations and initial planning period Euler equations) that can be estimated by Three Stage Least Squares (3SLS) in a rational expectations model. Computationally, this is a relatively easy approach compared to others discussed here, although it requires adjusting inputs to be quasi-fixed in the cost function and needs non-adjusting inputs to identify the cost function.

Prucha and Nadiri (1984) describe a numerical solution to the current and expected future cost minimisation problem, using a finite planning horizon (as opposed to the infinite planning horizons assumed by the other methods discussed so far) and certainty equivalence. For linear-quadratic technologies, certainty equivalence yields exactly the same factor input decisions as a closed-loop feedback based on a certain planning horizon. For more general technologies it can be regarded as a first-order approximation to a closed-loop feedback system. Prucha and Nadiri (1986) analyse and compare the statistical and computational efficiency of this approach (with the aid of a Monte Carlo study) in comparison to those of Epstein and Yatchew (1985) and Pindyck and Rotemberg (1983). They conclude that gains in statistical efficiency can be obtained by incorporating a full solution to the firm's cost minimisation problem (in contrast to Pindyck and Rotemberg who only minimise costs in the current planning period). Furthermore, they find that results from a finite planning horizon model can closely approximate those from an infinite planning horizon model, even when the planning horizons are moderate. Nadiri and Prucha (1989) present empirical work in the context of the telecommunications industry using both finite and infinite planning horizons. Prucha and Nadiri (1991) compare and analyse assumptions for these kinds of models in relation to the choice of finite or infinite time horizons, and continuous or discrete time solutions. However, it should be noted that all these methods are computationally very complex, reducing their attractiveness in empirical work.

Larson (1992) looks at adjustment cost models in the context of static and non-static expectations for technology advance, described in terms of an index that enhances the effectiveness of capital (ie. the higher the value of the index, the less the volume of capital that is required to produce a given level of output). Larson's paper illustrates the need to make explicit assumptions about the firm's intertemporal planning problems - in particular the form of production functions and the expectation forming processes - if a precise solution for factor demand equations is to be derived. This we do in our own work in this chapter.

Nickell (1986) presents a survey of dynamic models of labour demand and recommends a theoretically rigorous and empirically testable specification based on the introduction of a quadratic adjustment cost into the firm's net revenue maximisation problem. This is broadly the approach that we adopt in this chapter, but extending the analysis to include all the variable inputs.

For the second modification to the usual static model of the firm used in this chapter we draw from previous work on dynamic consumer demand systems. Consumer demand systems are analogous to factor demand systems in that they involve maximisation of a utility function (cf. minimisation of cost) subject to a budget constraint (cf. constraint of the production function). Solution of this problem gives demand as a function of income and prices (cf. output and prices in the case of the firm). Dynamic consumer demand systems are described by Spinnewyn (1981), Muellbauer and Pashardes (1988) and Pashardes (1986). The basic approach of all these papers is to express the utility maximisation problem not in terms of actual volumes of goods, but the services which they provide to the consumers, which may take account of the durability (asset life) and habit formation properties of each good and service. We can adopt an analogous approach in deriving dynamic factor demand and cost functions for firms. In doing this we take the services offered by the capital stock to be dependent on the depreciation rate which can be estimated in much the same way as the durability of goods in consumer demand systems.

We argue that these two modifications to the static framework (that is adjustment costs and unknown durability of the capital stock) are particularly important when seeking to model the banking sector. The inclusion of adjustment costs captures the transitory costs as companies switch to more technologically intensive methods of service provision. The estimation of a depreciation rate based on company behaviour allows for the measurement problem that follows from depreciation rates of plant and machinery used in preparing company accounts no longer being representative of the true asset lives of modern information technology equipment. The model which we derive is also applicable to other industries where technological change impacts on the fixed assets used in the production process.

Recently, researchers have attempted to apply static cost function models to financial service industries, and in particular to the banking sector. An excellent review of cost modelling of United States banks in Clark (1988) discusses different approaches used in defining inputs and outputs for the banking sector (which we shall draw on in later sections) and highlights issues that are encountered with the application of theory to this particular sector. The references quoted in this article should be supplemented by a number of more recent papers. Lawrence (1989) investigates generalising the usual transcendental logarithmic cost function using Box-Cox transformations, but finds this to be unnecessary. However, a very simple Cobb-Douglas specification of the cost function also fails to capture the innovation of computer technology in allowing inputs to be shared in a multi-product bank (ie. economies of scope). Noulas, Ray and Miller (1990) argue that most previous studies have been dominated by small institutions. They divide their sample of US banks into four size categories and model each separately. Finally, Ferrier and Lovell (1990) apply and compare both parametric models and Data Envelopment Analysis (DEA) to estimate technical and allocative efficiency in the banking industry. Their parametric model estimates a surprising degree of allocative inefficiency within the industry. The analysis in this chapter suggests that in part this may be due to the effect of adjustment costs in slowing down the convergence of factor inputs to their long-term levels.

In this chapter we are concerned primarily with the econometric estimation of efficiency, both in relation to inter-firm differences and gains in the sector as a whole over time. In the UK, some work on building societies has been done by Barnes (1985) in relation to efficiency gains resulting from mergers, by Hardwick (1989) in relation to economies of scale estimated from a transcendental logarithmic cost function, and by Drake and Weyman-Jones (1991 and 1992) in relation to scale, technical and allocative efficiency in a static Farrell type framework. However, no econometric modelling work has been attempted covering both banks and building societies, or have any estimates been made of efficiency gains over time. The comparison of banks and building societies is interesting since they now offer substantially similar services, whilst efficiency gains over time are interesting since they provide an indication of the benefits of the radical technological change that both sectors have experienced. It is this fact that makes a dynamic model incorporating adjustment costs particularly relevant to the banking sector.

This chapter will now proceed as follows. The next section gives a detailed statement of the basic dynamic model that will be used, section three introduces various adaptations to allow the model to be empirically applied to the UK banking sector (including the problem of inappropriate asset life assumptions used in published data). Section four describes the data that is used in the analysis, and presents the empirical results. Finally, section five presents the chapters principal conclusions.

2. A Model of Factor Demands with Adjustment Costs

Consider a firm that uses N inputs indexed $n=1,2,...,N$ to provide M outputs indexed $m=1,2,...,M$. Let $x^*_t=(x^*_{1t},x^*_{2t},...x^*_{Nt})$ denote the services offered by the input bundle in period t. For some inputs this may correspond to a routinely reported quantity (e.g. labour measured by the number of employees) but in other cases the link between services offered by a factor and quantities conventionally reported may not be so obvious (such as in the case of plant and machinery). We will return to this point in the next section. Let $w_t=(w_{1t},w_{2t},...,w_{Nt})$ denote the user costs associated with the factor services. The output bundle in period t is denoted $y_t=(y_{1t},y_{2t},...,y_{Mt})$. We assume that all elements of the vectors x^*_t, w_t and y_t are non-negative (ie. $x^*_t,w_t,y_t\geq 0$). We also assume that the firm is a price taker so that w_t is exogenous.

We assume that the cost of service provision for the firm in a given period t (net of any quasi-fixed inputs) is simply the sum of the factor payments in that time period. This can then be written as

$$A^*_t = \sum_{n=1}^{N} w_{nt} x^*_{nt} \tag{4.1}$$

We assume that production technology of the firm in period t is described by a transformation function

$$f(x^*_t,y_t,z_t,t) = 0 \tag{4.2}$$

where z_t is a vector denoting quasi-fixed inputs which for simplicity we assume has simply one element. The function $f(x^*_t,y_t,z_t,t)$ is assumed to be a differentiable and convex function of x^*_t. Note that the transformation function is allowed to change in each time period by virtue of the argument t. These shifts in the transformation function correspond to period by period changes in the total factor productivity of the firm, possibly as a result of technological change.

We now move on to suppose that the firm plans to provide services y_t at minimum cost and chooses x^*_t accordingly. We prefer to consider the objective of the firm to be cost minimisation, rather than profit maximisation for four reasons. First, under reasonable assumptions we can be sure that a solution always exists to the cost minimisation problem. Secondly, we intend to apply this model to building societies. Building societies are mutual institutions who do not seek to maximise profits, but do seek to provide a service for their members at minimum cost. Thirdly, we will later wish to define efficiency in terms of distance from a cost function. Lastly, an analysis based on the cost minimisation objective has the advantage of simplicity since it abstracts from the problem of determination of the scale of output of the firm.

The behaviour of the firm is, therefore, determined by minimising (4.1) subject to the constraint imposed by (4.2). Since there is no dependency of either cost or production between time periods, we are able to minimise the cost in each period independently of all others. The first order conditions that give the solution to this problem are easily derived by introducing Lagrange Multipliers μ^*_t as follows

$$w_{nt} + \mu^*_t \frac{\partial f}{\partial x^*_{nt}} = 0 \quad ; \quad n=1,2,...,N \tag{4.3}$$

$$f(x^*_t, y_t, z_t, t) = 0 \tag{4.4}$$

Equations (4.3) and (4.4) consists of N+1 equations that can be solved for x^*_t and μ^*_t to yield the optimum inputs of the company. It will be realised that in doing this we are now taking x^*_t to refer to the cost minimising input bundle.

The Lagrangian Multiplier can be solved by multiplying each equation in (4.3) by x^*_{nt} and summing to give

$$\mu_t^* = -A_t^* \Big/ x_t^* . \frac{\partial f}{\partial x_t^*} \tag{4.5}$$

The case of the firm we have just looked at is well known. However, we will now proceed to consider the case of a second firm which operates in identical conditions to the first with the exception of the fact that it faces additional adjustment costs in each period in which it alters the level of its inputs. Since this second firm will face a different cost structure, the optimum input bundle will be different and we will therefore denote it by $x_t=(x_{1t},x_{2t},\dots,x_{Nt})$ in period t. The outputs that the firm provides remain the same at y_t in period t, as does the quasi-fixed input z_t. The transformation function that the firm faces in this period remains the same, that is

$$f(x_t,y_t,z_t,t) = 0 \tag{4.6}$$

We now introduce a crucial difference between the first and second firm by the costs that are borne. For the second firm we assume that the cost of service provision in a given period is the sum of the factor payments in that time period, plus an adjustment cost that is related to the change in factor usage from the previous time period. We take the adjustment cost of an input to be proportional to its price and the square of the proportionate change in its volume of usage[13]. The cost to the firm incurred by service provision in period t (net of any quasi-fixed inputs) is, therefore, given by

$$a_t = \sum_{n=1}^{N} w_{nt} x_{nt} [1+\pi_n (x_{nt}-x_{nt-1})^2 / (2x_{nt-1} x_{nt})] \tag{4.7}$$

---

13 An alternative derivation follows from supposing that the effect of adjustments to a firm's factor inputs is to reduce their productivity. Following this line of thought we may specify the cost minimisation problem as to minimise $w_t.x_t$ subject to $f(x_t/\gamma_t,y_t)$ where γ_t is some suitable function involving changes in x_t. Depending on the choice of γ_t this will clearly result in a different model.

This is a different specification to the quadratic adjustment cost employed by Nickell (1986)[14]. Nickell's specification implies the property that if the firm expands along a path that increases all inputs by a fixed percentage in each period, then the adjustment cost will increase with the square of the inputs, thus taking up a progressively larger share of the total cost. We prefer a specification which has the property that if the firm expands along a path that increases all inputs proportionally, the adjustment cost remains a fixed proportion of total cost. We, therefore, have standardised the quadratic adjustment cost. This has the additional advantage of making the parameter π_n independent of the units of measurement of x_{nt}, thus easing its interpretation as the relative importance of adjustment costs compared to factor payments for each of the inputs. This interpretation is not true of the specification used by Nickell where a change in the units of measurement of x_{nt} would require a change in the value of π_n in order to avoid a change in the adjustment costs relative to the factor payments.

If adjustment costs are always positive, then $\pi_n > 0$ $(n=1,\dots,N)$.

We suppose that the firm considers its planning period costs which are calculated at time t by summing the discounted value of all future costs up to a planning horizon of T as follows:

$$A_t = \sum_{u=t}^{T} (1+r)^{t-u} a_u \tag{4.8}$$

$$= \sum_{u=t}^{T} (1+r)^{t-u} \sum_{n=1}^{N} w_{nu} x_{nu} \left[1+\pi_n (x_{nu}-x_{nu-1})^2/(2x_{nu-1}x_{nu})\right] \tag{4.9}$$

$$= \sum_{u=t}^{T} (1+r)^{t-u} \sum_{n=1}^{N} \left[w_{nu} x_{nu} + \pi_n w_{nu} x_{nu}^2/(2x_{nu-1}) - \pi_n w_{nu} x_{nu} + \pi_n w_{nu} x_{nu-1}/2\right]$$

where r is a discount rate. The process that generates the costs of

---

14 Nickell's specification of cost takes the form

$$\sum_{n=1}^{N} w'_{nt} x_{nt} [1+\pi_n (x_{nt}-x_{nt-1})^2]$$

the firm must take x_{t-1} as a fixed starting point.

If we suppose that the firm plans to provide services $y_t, y_{t+1}, \ldots, y_T$ at minimum planning period cost and chooses $x_t, x_{t-1}, \ldots, x_T$ accordingly, then the behaviour of this firm may be determined by minimising (4.9) subject to the constraint imposed by (4.6) and the initial conditions given by x_{t-1}. The solution to this problem will give the factor demands for the inputs to the second firm. They will differ from the case of the first firm by virtue of the adjustment costs providing a link between values of x_t in consecutive time periods.

The first order conditions of this problem are easily derived by introducing Lagrange Multipliers μ_u $(u=t,t+1,\ldots,T)$ as follows:

$$
\begin{aligned}
&(1+r)^{t-u} w_{nu}\left[1+\pi_n(x_{nu}/x_{nu-1}-1)\right]-(1+r)^{t-u-1}\pi_n w_{nu+1}\left[(x_{nu+1}/x_{nu})^2-1\right]/2 \\
&\quad +\mu_u\frac{\partial f}{\partial x_{nu}}=0 \quad ; \quad n=1,2,\ldots,N \quad ; \quad u=t,t+1,\ldots,T \qquad (4.10)
\end{aligned}
$$

$$f(x_u, y_u, z_u, u) = 0 \quad ; \quad u=t,t+1,\ldots,T$$

It will be realised that we are now taking x_u to refer to the cost minimising input bundle.

We can now see that this firm faces an additional problem in determining its optimal input requirements in period u; namely that it requires knowledge of x_{u+1} and w_{u+1}. We suppose the firm will form expectations of x_{u+1} and w_{u+1} which we shall write as $E(x_{u+1})$ and $E(w_{u+1})$, conditional on x_u and w_u respectively. If the firm only needs to plan factor demands for the current period, then we can restrict our attention to the case in equation (4.10) when $u=t$. Therefore, the behaviour of the firm in period t is found by the solution to the following equations:

$$w_{nt}[1+\pi_n(x_{nt}/x_{nt-1}-1)]-\pi_n E(w_{nt+1})[(E(x_{nt+1})/x_{nt})^2-1]/[2(1+r)]+\mu_t\frac{\partial f}{\partial x_{nt}}=0$$

$$n=1,2,\ldots,N \tag{4.11}$$

$$f(x_t,y_t,z_t,t) = 0$$

We now make assumptions about the way in which the firm forms expectations about the next period. Since the firm will be familiar with the phenomena of inflation in factor prices (e.g. the annual pay round), it seems reasonable to suppose that it will expect the price of factor n to rise by a proportion $\zeta_{1n}(1+r)$ in the next period. The inclusion of the discount rate in this factor recognises the tendency for periods of higher interest rates to correspond to periods of higher short-term inflation expectations. An expectation of the actual factor requirement is likely to be more problematic for the firm. However, we can note two properties that, in theory, we should anticipate this expectation to possess. First, we can note that quadratic adjustment costs penalise large changes, so the optimal response to a disturbance (caused by a change in output or factor prices) is a relatively large change in factors with low adjustment costs and a relatively small change in factors with high adjustment costs, followed by a sequence of smaller changes in the same direction. Following the initial adjustment, depending on whether x_t is rising or falling, we may expect x_{t+1} to be above or below x_t. The second property of the expected factor demands is that they should respond to factor price changes. Therefore, we may anticipate the expectation of x_{t+1} to depend on the expectation of w_{t+1} (through the own and cross-price elasticities on factor demands), and to be homogeneous of degree zero in its elements (ie. the same proportional change in <u>all</u> elements of w_{t+1} would not affect the expectation of x_{t+1}). Therefore, we can write the firm's expectations for period t+1 as follows:

$$E(w_{nt+1}) = \zeta_{1n}(1+r)w_{nt} \tag{4.12a}$$

$$E(x_{nt+1}) = \zeta_{2nt}x_{nt} \tag{4.12b}$$

where ζ_{2nt} is determined by the function

$$\zeta_{2nt} = \zeta_{2n}\left(x_{nt-1}/x_{nt-2}, E(w_{1t+1})/w_{1t}, E(w_{2t+1})/w_{2t}, \ldots, E(w_{Nt+1})/w_{Nt}\right)$$

$$= \zeta_{2n}\left(x_{nt-1}/x_{nt-2}, \zeta_{11}(1+r), \zeta_{12}(1+r), \ldots, \zeta_{1N}(1+r)\right) \qquad (4.12c)$$

such that $\zeta_{2n}\left(x_{nt-1}/x_{nt-2}, \zeta_{11}(1+r), \zeta_{12}(1+r), \ldots, \zeta_{1N}(1+r)\right)$ is homogeneous of degree zero in $(\zeta_{11}, \zeta_{12}, \ldots, \zeta_{1N})$ and

$$x_{nt-1}/x_{nt-2} < \zeta_{2nt} < 1 \qquad \text{if } x_{nt-1} < x_{nt-2}$$

$$1 < \zeta_{2nt} < x_{nt-1}/x_{nt-2} \qquad \text{if } x_{nt-1} > x_{nt-2}$$ [15]

Substituting these expectations (4.12a) and (4.12b) into equation (4.11) gives

$$w_{nt}\{\, 1 + \pi_n(x_{nt}/x_{nt-1} - 1) - \pi_n\zeta_{1n}(\zeta_{2nt}^2 - 1)/2 \,\} + \mu_t \frac{\partial f}{\partial x_{nt}} = 0 \qquad (4.13)$$

$n=1,2,\ldots,N$

In the interests of brevity we shall write $\zeta_{nt} = \zeta_{1n}(\zeta_{2nt}^2 - 1)/2$. This allows us to re-write (4.11) as

$$w_{nt}\{\, 1 + \pi_n(x_{nt}/x_{nt-1} - 1) - \pi_n\zeta_{nt} \,\} + \mu_t \frac{\partial f}{\partial x_{nt}} = 0 \qquad (4.14)$$

$n=1,2,\ldots,N$

---

15 The simplest specification that fulfills these requirements is

$$\zeta_{2nt} = (x_{nt-1}/x_{nt-2})^{\beta_n} + (1+r)\sum_{i=1}^{N}\alpha_{ni}\zeta_{1n} \text{ with } 0<\beta<1 \text{ and } \sum_{i=1}^{N}\alpha_{ni}=0$$

$$= \alpha_n + (x_{nt-1}/x_{nt-2})^{\beta_n}$$

where $\alpha_n = (1+r)\sum_{i=1}^{N}\alpha_{ni}\zeta_{1n}$ (a constant).

This was the empirical specification tested in the empirical work of this thesis.

As before the Lagrangian Multiplier can be solved by multiplying each equation in (4.14) by x_{nt} and summing to give

$$\mu_t = -\left(C_t + \sum_{n=1}^{N} \pi_n w_{nt} x_{nt} \{x_{nt}/x_{nt-1} - 1 - \zeta_{nt}\}\right) \Big/ x_t . \frac{\partial f}{\partial x_t} \tag{4.15}$$

where $C_t = \sum_{n=1}^{N} w_{nt} x_{nt}$ is the total factor payment in period t.

This can be simplified by writing

$$B_t = C_t + \sum_{n=1}^{N} \pi_n w_{nt} x_{nt} \{x_{nt}/x_{nt-1} - 1 - \zeta_{nt}\} \tag{4.16}$$

so that

$$\mu_t = B_t \Big/ x_t . \frac{\partial f}{\partial x_t} \tag{4.17}$$

At this stage it is worth reviewing the argument so far. We have formulated a model of factor demands for two firms. The first is similar to the traditional static model of the firm (although even here we have allowed for time related productivity changes), whilst the second is identical in every respect apart from additionally including adjustment costs associated with changes in factor usages. We now wish to go on to study the relationship between the behaviour of the two firms in order to find a practical method of distinguishing between them.

To make further progress we will need to make a restriction on the form of the transformation function; namely that

$$\left(\frac{\partial f}{\partial x_{nt}} \Big/ \frac{\partial f}{\partial x^*_{nt}} \right) = x^*_{nt}/x_{nt} \quad ; \quad n=1,2,\ldots,N \tag{4.18}$$

This assumption still caters for a wide variety of transformation functions[16].

We may now move on to investigate the more interesting relationships between the elements of x_t and x_t^* that emphasises the dynamic structure of x_t. These are more interesting since they allow us to directly identify how the behaviour of the two firms will differ.

Equations (4.14) and (4.3) can be combined to give

$$\mu_t \frac{\partial f}{\partial x_{nt}} \Big/ \mu_t^* \frac{\partial f}{\partial x_{nt}^*} = 1+\pi_n(x_{nt}/x_{nt-1}-1)-\pi_n\zeta_{nt}; n=1,2,\ldots,N \qquad (4.19)$$

Substituting for the two Lagrange multipliers from equations (4.17) and (4.5) gives

$$(B_t/A_t^*)\left(\frac{\partial f}{\partial x_{nt}} \Big/ \frac{\partial f}{\partial x_{nt}^*}\right)\left(x_t^* \cdot \frac{\partial f}{\partial x_t^*} \Big/ x_t \cdot \frac{\partial f}{\partial x_t}\right)=1+\pi_n(x_{nt}/x_{nt-1}-1)-\pi_n\zeta_{nt}$$

$$n=1,2,\ldots,N$$

and from (4.18)

$$(B_t/A_t^*)\left(\frac{\partial f}{\partial x_{nt}} \Big/ \frac{\partial f}{\partial x_{nt}^*}\right)= 1+\pi_n(x_{nt}/x_{nt-1}-1)-\pi_n\zeta_{nt} \quad ; \quad n=1,2,\ldots,N$$

and furthermore

$$(B_t/A_t^*)\ (x_{nt}^*/x_{nt}) =1+\pi_n(x_{nt}/x_{nt-1}-1)-\pi_n\zeta_{nt} \quad ; \quad n=1,2,\ldots,N \qquad (4.20)$$

---

16 An example of the transformation function for which this assumption would hold is

$$f(x_t,y_t,z_t,t) = h(y_t,z_t,t) - \sum_{n=1}^{N} \beta_n \log x_{nt} = 0$$

This is fairly general class of transformation function that includes the Cobb-Douglas as a special case when $h(y_t,z_t,t)=\log y_t$.

$$\Rightarrow \quad S^*_{nt} = w_{nt} x_{nt} \left[1 + \pi_n \left(x_{nt} / x_{nt-1} - 1\right) - \pi_n \zeta_{nt}\right] / B_t \quad ; \quad n = 1,2,\ldots,N \tag{4.21}$$

where $S^*_{nt} = \dfrac{w_{nt} x^*_{nt}}{A^*_t}$

An interesting interpretation of equation (4.20) follows from noting that $\left(x^*_{nt} / x_{nt}\right)$ is related to disequilibrium costs and $\pi_n \left(x_{nt} / x_{nt-1} - 1\right)$ is related to adjustment costs. $\pi_n \zeta_{nt}$ enters the equation because of the importance of expectations in minimising future adjustment costs. Equation (4.20) is also analogous to Treadway's (1974) flexible accelerator model. Non-theory based studies often take the relationship between $\left(x^*_{nt} / x_{nt}\right)$ and $\left(x_{nt} / x_{nt-1}\right)$ to be linear, that is, a simple partial adjustment model. However, equation (4.20) shows that this is not the case (as already illustrated by Larson (1992)).

Under a wide variety of transformation functions, and in particular that of footnote 16, we can show (see Appendix A) that[17]

$$A^*_t = A_t \prod_{n=1}^{N} (S_{nt}/S^*_{nt})^{S^*_{nt}} \tag{4.22}$$

It is easily verified that $A_t \geq A^*_t$. We would expect this to be the case

---

[17] An alternative formulae would follow from the approach of Schmidt (1984), and developed further by Kumbhakar (1991). This involves writing

$$S_{nt} = S^*_{nt} + U_{nt} \quad ; \quad n=1,2,\ldots,N$$

where to ensure that input factor shares sum to one we must have $\Sigma U_{nt}=0$. In the case of transcendental logarithmic cost functions we can apply Shephard's Lemma to both $\log A_t$ and $\log A^*_t$ to get

$$\frac{\partial \log A_t}{\partial \log w_{nt}} = S_{nt} \quad ; \quad n=1,2,\ldots,N$$

and $$\frac{\partial \log A^*_t}{\partial \log w_{nt}} = S^*_{nt} \quad ; \quad n=1,2,\ldots,N$$

so that

$$\frac{\partial \log(A_t/A^*_t)}{\partial \log w_{nt}} = U_{nt} \quad ; \quad n=1,2,\ldots,N$$

Kumbhakar (1991) shows that these requirements are satisfied if

$$\log(A_t/A^*_t) = u'_t E^{-1}_t u_t/2$$

where $u_t=(U_{1t},U_{2t},\ldots,U_{n-1t})$ and E_t is the matrix of price coefficients in the transcendental logarithmic function with the nth row and column deleted (so that it is not singular).

since adjustment costs act as incentive for firms to adopt factor demands that result in a greater cost than would be the case if there were no adjustment costs. Given fixed B_t and A^*_t, (4.20) can be written as a quadratic equation in x_{nt}.

$$(\pi_n/x_{nt-1})x^2_{nt} + (1-\pi_n-\pi_n\zeta_{nt})x_{nt} - (B_t/A^*_t)x^*_{nt} = 0 \; ; \; n=1,2,...N \tag{4.23}$$

If $\pi_n=0$ (ie. no adjustment costs), then a solution to these equations is provided by $x_{nt}=x^*_{nt}$. Otherwise we can write

$$x_{nt}=\{\pi_n+\pi_n\zeta_{nt}-1\pm[(1-\pi_n-\pi_n\zeta_{nt})^2+4\pi_n(B_t/A^*_t)(x^*_{nt}/x_{nt-1})]^{1/2}\}x_{nt-1}/2\pi_n$$

We have here two solutions. However, it is easily verified that since $A^*_t,x_{nt-1},x^*_{nt},\pi_n>0$ and B_t will only be negative in periods of rapidly declining output, only the solution with the positive sign is likely to yield a positive value for x_{nt}. Therefore

$$x_{nt}/x_{nt-1}=\{\pi_n+\pi_n\zeta_{nt}-1+[(1-\pi_n-\pi_n\zeta_{nt})^2+4\pi_n(B_t/A^*_t)(x^*_{nt}/x_{nt-1})]^{1/2}\}/2\pi_n$$

For empirical work we will find it useful to express this equation in terms of factor shares,

$$x_{nt}/x_{nt-1}=\{\pi_n+\pi_n\zeta_{nt}-1+[(1-\pi_n-\pi_n\zeta_{nt})^2+4\pi_n S^*_{nt}B_t/(w_{nt}x_{nt-1})]^{1/2}\}/2\pi_n \tag{4.24}$$

$$n=1,2,...,N$$

It is helpful to understand the implication of this equation for the relationship between the behaviour of the two firms in economic terms. We have already noted that when $\pi_n=0$, then $x_{nt}=x^*_{nt}$. In other words, x^*_{nt} is the value that x_{nt} will take (regardless of x_{nt-1}) if there are no costs associated with adjustments in factor demands in period t. Furthermore, it can easily be seen that when w_{nt} and all other exogenous variables stay fixed and $\zeta_n=0$ ($n=1,2,...N$), then equations (4.24) will converge to $x_{nt}=x^*_{nt}$ as $t\to\infty$. In this sense we can consider x^*_{nt} to be the "ideal" level of x_{nt}.

To further understand equation (4.24) it is helpful to consider the special case when $x_{nt-1}=x^*_{nt}$ (ie. in period t-1 the firm is already in its ideal position for period t with regard to input factor n) and $\zeta_{nt}=0$ ($n=1,2,\ldots,N$). Here we find that $x_{nt}=x_{nt-1}$ (ie. there is no change in factor demands). The higher is the ratio x^*_{nt}/x_{nt-1}, the higher is x_{nt}/x_{nt-1} (ie. the greater the distance between a factor demand in period t-1 and its ideal level in period t, the greater will be the adjustment to that factor in period t).

3. Empirical Specification

In this section we make the extensions we require to the model in order to apply it to empirical work. These concern two areas: first, the problems of measurement of the stock of plant and machinery, and second, the problem that S^*_{nt} is not directly observed.

We begin by addressing the problem that we may not be able to directly measure factor services (x_{nt}) and their user cost. In particular, this will be the case for plant and machinery where the book value has been computed using some arbitrary depreciation rate. Although this depreciation rate may at one time have been appropriate, it is possible that with changes in technology it now bears little relevance to the useful asset lives, and results in an incorrect estimate of capital stock and an under-estimate of the user cost. However, we may specify a relationship between the measured factor quantities and the actual factor services as follows.

Let $x^+_t=(x^+_{1t},x^+_{2t},\ldots,x^+_{Nt})$ denote the input bundle according to the firm's conventional accounting practice (using GBV to measure plant and machinery input) in period t and let the corresponding input user-price vector be denoted by $w^+_t=(w^+_{1t},w^+_{2t},\ldots,w^+_{Nt})$. Let the depreciation rate for input factor n assumed in the firm's accounts be δ_{1n} and let the $(n\times n)$ diagonal matrix containing these rates be Δ_1. Let the actual depreciation rate be δ_{2n} and let the $(n\times n)$ diagonal matrix containing these rates be Δ_2. Then expenditure on plant and machinery in period s can be expressed in terms of either measured GBV or factor services (taken as being useful factor stock) as follows

$$x^+_t - \Delta_1 x^+_{t-1} = x_t - \Delta_2 x_{t-1} \tag{4.25}$$

Re-writing in terms of growth rates

$$x^+_{nt}/x^+_{nt-1} = (x_{nt-1}/x^+_{nt-1})[x_{nt}/x_{nt-1} - \delta_{2n}] + \delta_{1n} \quad ; \quad n=1,2,\ldots,N$$

Substituting from equation (4.24) gives

$$x^+_{nt}/x^+_{nt-1} = (x_{nt-1}/x^+_{nt-1}) \\ (\{\pi_n + \pi_n\zeta_{nt} - 1 + [(1-\pi_n - \pi_n\zeta_{nt})^2 + 4\pi_n S^*_{nt} B_t/(w_{nt} x_{nt-1})]^{1/2}\}/2\pi_n - \delta_{2n}) + \delta_{1n} \\ n=1,2,\ldots,N \tag{4.26a}$$

The right hand side of this equation now contains only past values of x_{nt} and x^+_{nt} (that is x_{nt-1} and x^+_{nt-1}). x_{nt-1} can be constructed over time from the relationship

$$x_{t-1} = \Delta_2 x_{t-2} + x^+_{t-1} - \Delta_2 x^+_{t-2} \tag{4.26b}$$

after assuming starting values, $x_{n1} = x^+_{n1}(1-\delta_{1n})/(1-\delta_{2n})$ ($n=1,2,\ldots,N$).

The relationship between the price of a unit of x^+_{nt} and x_{nt} can be shown to be (see Appendix B)

$$w^+_{nt} = \frac{1-\delta_{1n}\zeta_{1n}}{1-\delta_{2n}\zeta_{1n}} w_{nt} \tag{4.27}$$

We now need to specify how we believe the long-run steady state factor shares (S^*_{nt}) are determined. This is essential since S^*_{nt} is not directly observed. We opt to work with the factor shares (in preference to factor demands), for two principal reasons. First, as we shall soon see, this allows x^*_{nt} to be specified in a theoretically satisfactory manner (ie. so as to be consistent with requirements of homogeneity and symmetry), and secondly this alleviates the problem of heteroscedasticity during econometric estimation.

We suppose the underlying cost function of the firm to take a transcendental logarithmic form popularised by Jorgenson and Lau (1977). This function is a flexible form that will provide a second order approximation to any arbitrary cost function, and is widely used in empirical research. Dummy variables have been introduced for each firm (d_{1s}) and time period (d_{2t}). The dummy variables for individual firms allow for the various characteristics of the firms that cannot be picked up by other variables. These will include permanent differences in efficiency between firms. Efficiency changes over time are incorporated into the model by the dummy variables for individual time periods. This treatment of firm specific and time specific efficiency is the same as in Bartelsman and Dhrymes (1991). Certain second order terms were omitted when they were found to be consistently insignificant at the 5% level in trial estimations of the model. This is especially desirable since a number of recent authors have noted the lack of robustness in full transcendental logarithmic specifications. In particular, Roller (1990) notes what he calls the "flip-flop" property of the transcendental logarithmic function, whereby small changes in the second order terms from negative to positive may cause the shape of the cost function surface to invert. This can have dramatic effects on efficiency when measured in terms of deviations from the cost function. In the light of this finding we were very cautious about the inclusion of too many second order terms in the function (especially when there is a risk of multicollinearity in their estimation). We write the underlying cost function for firm s as

$$\begin{aligned}
\log A_t^* &= \upsilon + \upsilon_s d_{1s} + \phi_t d_{2t} \\
&+ \sum_{n=1}^{N} \alpha_n \log w_{nt}^+ + 1/2 \sum_{n=1}^{N} \sum_{i=1}^{N} \alpha_{ni} \log w_{nt}^+ \log w_{it}^+ + \sum_{n=1}^{N} \upsilon_{ns} d_{1n} \log w_{nt}^+ \\
&+ \sum_{m=1}^{M} \beta_m \log y_{mt} + 1/2 \sum_{m=1}^{M} \sum_{j=1}^{M} \beta_{mj} \log y_{mt} \log y_{jt} + \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma_{mn} \log w_{nt}^+ \log y_{mt} \\
&+ \chi_1 \log z_t + 1/2 \ \chi_2 (\log z_t)^2 + \sum_{n=1}^{N} \chi_{3n} \log w_{nt}^+ \log z_t + \sum_{m=1}^{M} \chi_{4m} \log y_{mt} \log z_t \\
&+ \sum_{n=1}^{N} \tau_n t \log w_{nt}^+
\end{aligned} \tag{4.28a}$$

$$\text{where } \sum_{n=1}^{N} \alpha_n = 1$$

$$\sum_{n=1}^{N} \alpha_{ni} = 0 \quad ; \quad i=1,2,\ldots,N$$

$$\alpha_{ni} = \alpha_{in} \quad ; \quad i,n=1,2,\ldots,N$$

$$\sum_{n=1}^{N} \upsilon_{sn} = 0 \tag{4.28b}$$

$$\sum_{n=1}^{N} \gamma_{nm} = 0 \quad ; \quad m=1,2,\ldots,M$$

$$\sum_{n=1}^{N} \chi_{3n} = 0$$

$$\sum_{n=1}^{N} \tau_n = 0$$

so as to ensure symmetry of the cross-price effects in the demand functions and homogeneity in factor prices.

A restriction of global constant returns to scale would also imply

$$\sum_{m=1}^{M} \beta_m = 1$$

$$\sum_{m=1}^{M} \sum_{j=1}^{M} \beta_{mj} = 0 \tag{4.28c}$$

Shephard's Lemma can be used to show that for firm s

$$S^*_{nt} = \alpha_n + \sum_{i=1}^{N} \alpha_{ni} \log w^+_{it} + \sum_{m=1}^{M} \gamma_{nm} \log y_{mt} + \chi_{3n} \log z_t + \tau_n t + \upsilon_{sn} d_{1n}$$

$$n=1,2,\ldots,N \tag{4.29}$$

Interpretations can be placed on many of the coefficients in this transcendental logarithmic cost function. This is made intuitively easier if we add a constant to all variables so that they have a zero mean.

υ_s and ϕ_t are fixed effects which, to the extent that they are not captured by other variables, model the efficiency effects pertinent to a particular firm and a particular time period, relative to the reference firm in the reference time period.

α_n gives the underlying factor share of input n for the reference firm when all variables are at their means.

α_{ni} determines the underlying price elasticity of input n with respect to the price of input i through the formulae

$$\eta_{ni} = \alpha_{ni}/S^*_{ni} + S^*_{ni} \quad ; \quad \text{if } n \neq i \tag{4.30a}$$

$$\eta_{nn} = \alpha_{nn}/S^*_{nn} + S^*_{nn} - 1 \quad ; \quad \text{if } n = i \tag{4.30b}$$

υ_{sn} gives the difference from the reference firm of the firm s factor share of input n (as a result of different methods of production adopted by the firm and allocative inefficiency).

β_m and β_{mj} model the way in which cost is affected by the level of output, and γ_{mn} models the impact of changes in the level of output m on the factor share of input n.

χ_1 and χ_2 model the way in which cost is affected by the level of the quasi-fixed input, and χ_{3n} models the impact of changes in the level of the quasi-fixed input on the factor share of input n.

Finally, τ_n is a time trend in the factor share of input n, reflecting changes in the methods of production by the firms.

4. Empirical Application: UK Banking Sector

Estimated parametric models of the banking and building society sectors have been constructed before. For example, Hadjimatheou (1976) built a sectoral econometric model covering the markets for new housing and mortgage financing by building societies. The equations for the latter dealt with the demand and supply of mortgage advances. Our cost function system takes advances and other measures of bank and building society "outputs" as exogenous variables and estimates the relationships determining the factor demands (labour, and plant and equipment) and costs, and the cost efficiency of providing the outputs.

a. *Data*

Empirical modelling required the construction of a database containing inputs and outputs for a sample of banks and building societies over a number of years. The principal source of this data was company reports and accounts and annual returns to the Building Societies Association (BSA). Selection of banks and building societies to include in the sample was mainly governed by the availability of sufficient time series data. The sample contained 7 banks (Barclays, National Westminster, Midland, Lloyds, Bank of Scotland, Royal Bank of Scotland and Standard Chartered) and 4 building societies (Halifax, National Provincial, Bristol and West and the Chelsea) giving a total of 114 observations. In the interests of brevity we will refer to both types of firms as "banks". Details of data sources are given in the Annex at the end of this thesis.

Output of a bank can not usually be measured directly. Clark (1988) identifies two approaches that can be adopted. The first he refers to as the intermediation approach. Banks are viewed as collecting deposits and purchasing funds to be subsequently intermediated into loans and other assets. Thus deposits are treated as inputs along with capital and labour, and only dollar volumes of assets are treated as outputs. The second approach is referred to by Clark as the production approach. Here banks are producers of services associated with individual loan and deposit accounts. These accounts are "produced" using capital and labour.

The choice between the two approaches reflects the definition of output (and hence efficiency) that we wish to adopt. For this thesis we are interested in studying how the banks use their physical resources (labour and equipment) to generate the different kinds of transactions that serve customers (whether by deposit or loan facilities). Therefore, we adopt a production approach and attempt to find variables that will proxy the unobserved level of transactions that constitute the services of deposit and loan accounts. Financial data is available from annual reports and accounts giving the total value of deposits and advances from individual banks and building societies. However, banks and building societies offer different categories of accounts distinguished by the level of service provided and the interest paid. Thus, a current account offers the customer a range of facilities (e.g. immediate withdrawal, standing orders, etc.) at the expense of a low rate of interest, whilst a deposit account typically offers little service but pays a high rate of interest. This inverse relationship between the level of service provided and the interest paid enables us to use the inverse of interest payments as a proxy variable for the average level of service provided by the bank to its depositors. Typically this will tend to be lower where a bank specialises in deposit rather than current accounts. The importance of distinguishing between deposit and current accounts is confirmed by other researchers (e.g. Ferrier and Lovell (1990)).

Three principle input variables were considered: labour, plant and machinery, and buildings. Details of data sources are given in the Annex 1 at the end of this thesis.

b. *Results*

Equations (4.26), (4.28) and (4.29) form the basis for the empirical work in this section, taking labour and plant and machinery to be two variable inputs and buildings to be a quasi-fixed input (similar to Hardwick (1989)). Efficiency of the individual banks and the sector as a whole in individual years is measured by two sets of dummy variables placed in the cost function. The size of each dummy variable measures the scalar by which the cost function for that particular bank, or for that year, must be reduced in order to coincide with the cost function for the reference bank in the reference year (Bank (1) in 1987).

A two stage estimation procedure was used. In the first stage of the estimation each of the dynamic factor demand equations (4.26) (with S^*_{nt} substituted from (4.29)) were estimated independently by non-linear least squares (employing a Gauss-Newton algorithm), imposing all intra-equation parameter restrictions in (4.28b)[18]. Estimates of S^*_{nt}, A^*_t and the adjusted user cost of plant and machinery were then constructed using equations (4.21), (4.22) and (4.27) respectively.

In the second stage of the model estimation the non-linear parameters (the actual plant and machinery depreciation rate and the adjustment cost coefficients) were imposed at their values estimated in the first stage, and the system of factor share equations (4.29) and the cost function (4.28) were estimated as Seemingly Unrelated Regressions (SUR) imposing both intra and inter-equation restrictions in (4.28b). (One factor share equation - the labour equation - being omitted in order to avoid over-identification of the system.) Under the usual regression model assumptions, this procedure gives unbiased and asymptotically efficient estimates of all parameters.

At an early stage in the data analysis a specification was tested for the function $\zeta_{2nt}=\zeta_{2n}\left(x_{nt-1}/x_{nt-2},\zeta_{11}(1+r),\zeta_{12}(1+r),\ldots,\zeta_{1N}(1+r)\right)$ as shown in footnote 15. However, since the hypothesis that $\alpha=1$ and $\beta=0$ (implying $\zeta_{2nt}=1$) and $\zeta_{1n}=0$ could be accepted at the 5% level of significance, and inclusion of these extra parameters excessively complicated the model, we judged that it was appropriate to set $\zeta_{2nt}=1$ and $\zeta_{1n}=0$. This suggests that the banks in the sample generally expected to retain the same level of factor inputs and that factor prices would rise in line with discount rates.

Trial model estimations using separate data on banks and building societies indicated that many of the structural coefficients in the

---

[18] The principal econometric problem in the first stage of the estimation is that each equation includes the variable B_t which contains an endogenous component in its summation. This was avoided by replacing (x_{nt}/x_{nt-1}) in B_t by its value in the previous period, (x_{nt-1}/x_{nt-2}). In the second stage of the estimation procedure this problem does not exist.

cost function differ between the two groups of banks. In order that the model may provide an adequate benchmark against which to measure efficiency within both sectors a number of dummy variables where introduced relating to the output and branch networks of just the building societies. These capture the differences between the sectors with respect to these two variables, reflecting differing organisational structures and attitudes as to the role of branch networks.

Trial model estimates also indicated that whereas banks are subject to increasing returns to scale, building societies are subject to decreasing returns to scale (consistent with the general findings of Hardwick (1989), and Drake and Weyman-Jones (1991) and (1992)). This was verified by a formal statistical test that indicated the hypothesis of global constant returns to scale, implied by the restrictions (28c), could be rejected at the 5% level of significance. However, since we are interested in measuring efficiency (including scale efficiency) a model that imposed constant returns to scale was used in the analysis[19].

Table I presents the estimated coefficients (excluding dummy variables which are shown later) and Table II presents a range of test statistics. The adding up and homogeneity restrictions can not be accepted at the 5% level of significance, however, this is not unusual with this type of model.

---

19 Since the estimated cost function includes a dummy variable for each bank, the degree of non-constant returns to scale must be identified by changes in the scale of the output of each bank over time. It is possible that only the short term returns to scale are identified by the coefficients on output in the cost function, leaving the longer term effects to be picked up by the "efficiency dummy variables". If this is the case, interpretation of both the estimated returns to scale and efficiency becomes very difficult. This confusion is avoided if constant returns to scale are imposed in the model which is used to estimate efficiency (efficiency then being defined to include scale efficiencies).

	Coefficient	Std. Error	Building Society Dummy Coefficient	Std. Error
υ	5.0393	0.0248		
υ_1	-0.1019	0.0200		
υ_2	0.0476	0.0202		
υ_3	-0.2316	0.0197		
υ_4	-0.0541	0.0222		
υ_5	-0.0039	0.0224		
υ_6	-1.4978	0.0218		
υ_7			-1.2100	0.0482
υ_8			-1.2663	0.0523
υ_9			-0.9041	0.0454
υ_{10}			-1.0509	0.0361
ϕ_{78}	0.4650	0.0392	-0.0908	0.0553
ϕ_{79}	0.4543	0.0283	-0.0168	0.0439
ϕ_{80}	0.3947	0.0280	-0.1613	0.0433
ϕ_{81}	0.2636	0.0305	-0.1474	0.0458
ϕ_{82}	0.1590	0.0309	0.0675	0.0442
ϕ_{83}	0.1858	0.0289	-0.0323	0.0418
ϕ_{84}	0.0602	0.0282	-0.0016	0.0408
ϕ_{85}	0.0796	0.0274	0.0297	0.0397
ϕ_{86}	0.0861	0.0266	-0.0493	0.0392
α_k	0.2192	0.0195		
α_{kk}	-0.1051	0.0716		

Table I: Coefficient Estimates

	Coefficient	Std. Error	Building Society Dummy Coefficient	Std. Error
υ_{k1}	-0.3048	0.0781		
υ_{k2}	-0.2078	0.0562		
υ_{k3}	-0.2368	0.0754		
υ_{k4}	-0.1939	0.0677		
υ_{k5}	-0.0013	0.0396		
υ_{k6}	0.0877	0.0453		
υ_{k7}	0.1751	0.0768		
υ_{k8}	0.2748	0.1208		
υ_{k9}	-0.2136	0.0466		
υ_{k10}	0.0530	0.0511		
β_y	0.7411	0.0950		
β_i	-0.3568	0.0736	-0.5818	0.0970
β_a	0.6157	0.0565	0.1113	0.0901
χ_l			0.4705	0.0538
γ_{yk}	0.1051	0.0389		
τ_k	-0.0019	0.0049		
π_l	6.5035	2.1373		
π_k	1.8688	0.7703		
δ_k	0.7473	0.0211		

Observations:	Factor shares	114
	Cost function	114
Estimated coefficients: (including dummy vars.)	Factor shares	27
		-1 intra equation restrictions
	Cost function	67
		-17 intra equation restriction
Additional inter equation restrictions:		-12

Standard Error: factor shares	0.0704
Standard Error: cost function	0.0920
R squared: factor shares	0.7962
R squared: cost function	0.9991

Table I: Coefficient Estimates (*continued*)

Test	Distribution	Test Statistic
Homogeneity and Adding-up	$Chi^2(29)$	121.183
Constant Returns to Scale	$F(2,162)$	5.731
AR(1)		
Factor Share Equations	$Chi^2(1)$	1.404
Cost Function	$Chi^2(1)$	2.628
Whole System*	$Chi^2(2)$	1.643
ARCH(1)		
Factor Share Equations	$Chi^2(1)$	0.570
Cost Function	$Chi^2(1)$	2.213
Whole System*	$Chi^2(2)$	3.768
Structural Stability		
Banks vs. Building Societies		
Factor Share Equations**	$F(3,85)$	0.591
Cost Function**	$F(3,60)$	1.376
Whole System	$F(4,160)$	1.793
Data from (1978-1987) vs. data (1978-1986)		
Factor Share Equations**	$F(9,79)$	1.382
Cost Function**	$F(8,55)$	0.444
Whole System	$F(17,147)$	0.960
Joint Significance of Bank Dummy Variables $(v_1,v_2,\ldots,v_{10},v_{k1},v_{k2},\ldots,v_{k10},v_{l1},v_{l2},\ldots,v_{l10})$	$F(20,164)$	149.359
Joint Significance of all Time Dummy Variables $(\varphi_{78},\varphi_{79},\ldots,\varphi_{86},\varphi_{s76},\varphi_{s77},\ldots,\varphi_{s86})$	$F(19,164)$	35.176
Joint Significance of Building Society Time Dummy Variables $(\varphi_{s76},\varphi_{s77},\ldots,\varphi_{s86})$	$F(11,164)$	1.832

* These tests are constructed from the hypothesis that $(b_s,b_c)=0$; where $u_{st}=X_{st}b+b_s u_{st-1}+e_{st}$, $u_{ct}=X_{ct}+b_c u_{ct-1}+e_{ct}$ (in the case of AR(1)) or $u^2_{st}=b_s u^2_{st-1}+e_{st}$, $u^2_{ct}=b_c u^2_{ct-1}+e_{ct}$ (in the case of ARCH(1)), estimated by SUR; and u_{st} and u_{ct} are the estimated residuals in the factor share equation and cost function respectively.

** The degrees of freedom associated with the tests on individual equations take no account of the additional degrees of freedom resulting from the cross-equation constraints.

Table II: Summary of Test Statistics

The overall performance of the estimated model is good. There is no evidence of serial correlation in any of the equations (at a 5% level of significance), however, we do note evidence of heteroscedasticity in the cost function equation (although still not significant at the 5% level)[20]. This would compromise the statistical efficiency of the estimates but will not introduce any bias.

There are a number of interesting features of the equations that need to be discussed.

Beginning with the underlying cost function and factor shares, it is interesting to note the implied price elasticities of the factor demands. These can be calculated from equation (4.30) for the underlying long-term levels of factor demand (ie. after full adjustment) with respect to the wage rate and the user cost of plant and machinery. Equation (4.30) shows the price elasticities vary with the factor shares. When computed at the mean level of factor shares over the sample period the own-price elasticity of labour is -0.0626 and the own-price elasticity of plant and machinery is -0.2622. These are low, indicating that in this sector factor demands are inelastic with respect to input prices.

Turning to the factor share equations, the positive coefficient on income in the plant and machinery equation indicates that there has been a tendency for banks to become more equipment intensive as output increases (ceteris paribus).

It is also interesting, although not altogether surprising, that the econometric results indicate a faster depreciation rate is used by the banks in their investment decisions than is generally used in their

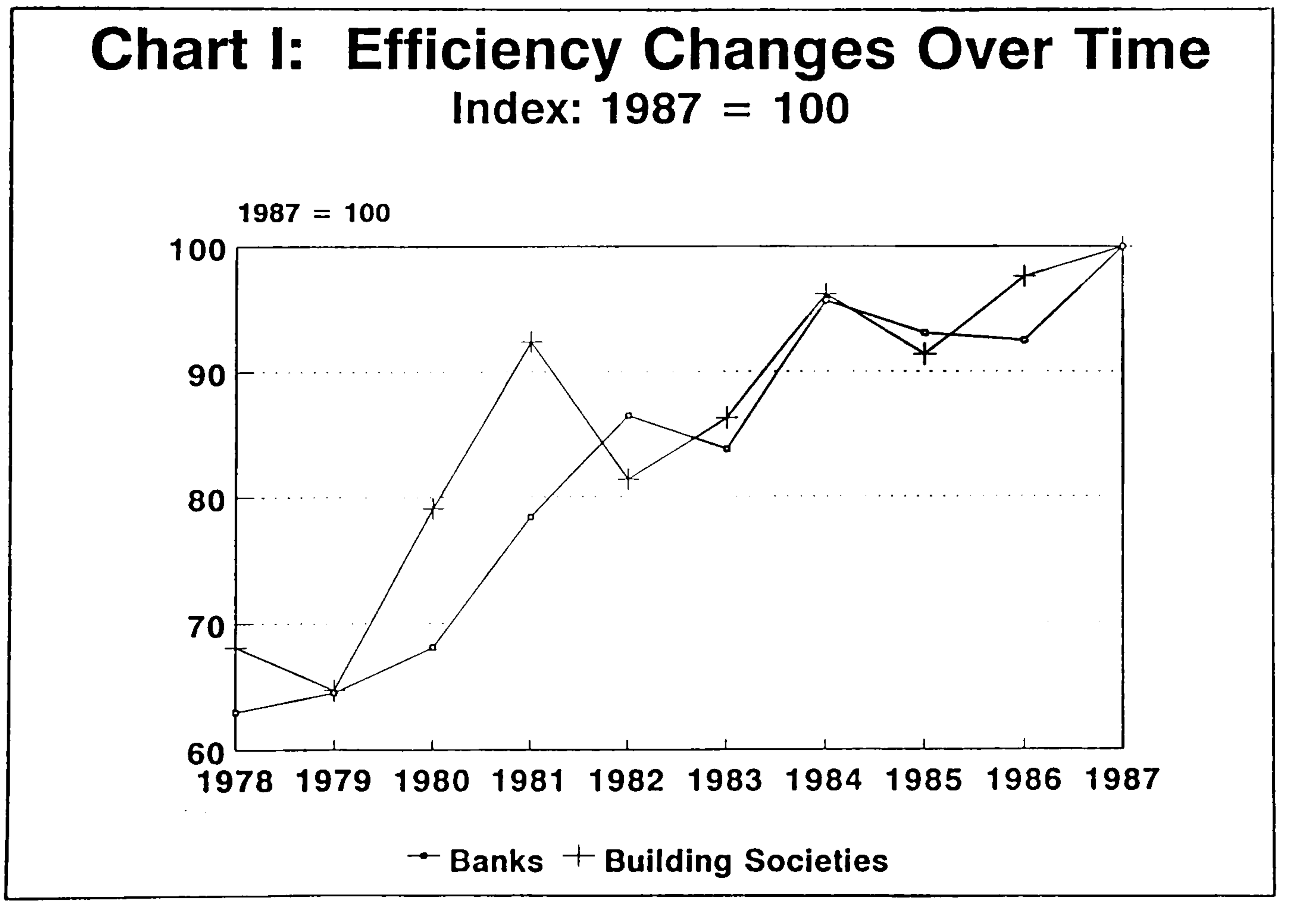
---

20 Heteroscedasticity in the pooled residuals from banks and building societies was found in chapter III. This was dealt with by estimating the pooled model by Generalised Least Squares (GLS) which modelled a different residual variance for each of the two groups.

accounting procedures. The estimated value is 0.7473, compared to 0.8 generally used in industry accounts. The difference between these figures is significant at the 5% level (using 't' distribution for the estimated depreciation rate). This most probably reflects the fact that accounting practices lag behind the changing technology of the industry.

The next area for discussion is the magnitude of the factor adjustment costs in the banking sector. The model predicts that a 1% increase in the demand for labour will result in a small 0.07% rise in the cost of each unit of labour in the year of the increase. However, because of the quadratic nature of the adjustment cost, a 10% increase in the demand for labour will result in a much more noticeable 6.5% rise in the cost of each unit of labour. These adjustment costs may reflect additional training needs associated with recruiting a large number of new staff, but also may be partly picking up an effect from an upward sloping short-run demand curve for the type of labour employed by banks (although they may still be a price-takers in the long-term). Turning now to plant and machinery, the adjustment costs appear to be smaller. The model predicts that a 1% increase in the stock of plant and machinery will result in an almost imperceptible 0.02% rise in the user cost in the year of the increase. However, a 10% increase in the stock of plant and machinery will result in a 1.9% rise in the user cost of plant and machinery. This most likely represents the cost of installation and lower productivity before employees are familiar with the new equipment.

Finally, we come to consider the efficiency effects estimated by the model. First, we note from Table II that both sets of dummy variables (for individual banks and for individual years) are significant at the 5% level. Table III and the associated Chart I display the implied efficiency indices for the UK banking sector over the period 1978 to 1987 as measured by the dummy variables for each year in the cost function. Efficiency rose only slowly between 1978 and 1979. Between 1979 and 1984 substantial gains were made by both sectors. This co-incided with a series of developments in the banking and building society sectors which have loosened restrictions on lending activities and increased competitive pressures. The most important are the ending of guidelines on building society lending (1979), the abolition of the so-called *corset* Supplementary Special Deposits Scheme designed to curb bank lending (1980), the abolition of the Reserve Asset Ratio requirement under which banks had to hold 12.5% of their deposits in a specified range of liquid assets (1981), the abolition of hire-purchase restrictions (1982) and the collapse of the building society cartel (1983). It is to be expected that the freeing of these restrictions, and the spurs to competitive pressure, would result in some growth of efficiency in the sector. Since 1984, efficiency in the two sectors appears to have grown in tandem at a slower rate. The average annual rise in efficiency over the period 1978 to 1987 has been 5.3% for banks and 4.4% for building societies. Both of these estimates are higher than estimates of total factor productivity for the UK economy as a whole over this period.



Year	Banks Efficiency Index	Building Societies Efficiency Index
1978	63.0	68.1
1979	64.5	64.7
1980	68.1	79.1
1981	78.4	92.4
1982	86.5	81.4
1983	83.8	86.3
1984	95.7	96.2
1985	93.1	91.4
1986	92.5	97.6
1987	100.0	100.0

Table III: Efficiency Estimates for the UK Banking Sector

Bank/ Building Society	Efficiency Index (rank order in parenthesis)	
Banks		
Bank (1)	79.3	(5)
Bank (2)	87.8	(2)
Bank (3)	75.6	(6)
Bank (4)	100.0	(1)
Bank (5)	83.7	(3)
Bank (6)	79.6	(4)
Standard Deviation	8.7	
Building Societies		
Society (1)	94.5	(2)
Society (2)	100.0	(1)
Society (3)	69.6	(4)
Society (4)	80.6	(3)
Standard Deviation	13.7	

Table IV: Efficiency Estimates for the UK Banking Sector

Table IV shows the results for individual banks and societies in the sample. It is immediately clear from the estimates of the dummy variables in Table I that there are three groups of banks: Banks (1) to (6) (retail banks), Bank (7) (Standard Chartered Bank - a specialist commercial bank) and building societies. Because of the different nature of the businesses of these three groups the results are not comparable. Therefore, a separate efficiency index has been calculated for each group, based on the most efficient bank of the group as 100.0. The most obvious feature of these results is the considerably greater degree of variation in the efficiency performance of the individual building societies (a range of 100.0 to 69.6 with a standard deviation of 13.7) compared to the banks (a range of 100.0 to 75.6 with a standard deviation of only 8.7).

Table V shows the relative size of adjustment costs faced by the banking sector over the data period. These are generally highest for equipment (reflecting faster growth in usage of this factor compared to labour). Adjustment costs for equipment are highest in 1981 and 1982 - corresponding to the period of entry of banks into the mortgage market following the abolition of the *Corset* credit controls in 1980, and 1985 and 1986 - corresponding to the lead up to *Big Bang* and financial service liberalisation in 1986.

Year	Adjustment costs for each factor as % of labour and equipment factor payments Labour	 Equipment	 Total
1978	1.9	4.2	2.5
1979	1.9	4.2	2.4
1980	2.8	3.2	2.9
1981	1.4	4.5	1.9
1982	1.5	5.6	2.4
1983	0.9	1.8	1.1
1984	1.3	2.0	1.5
1985	1.8	6.9	3.8
1986	2.9	4.3	3.2
1987	2.8	1.4	2.5

Table V: Average Adjustment Costs Estimates for the UK Banking Sector

Year	Increase in labour and equipment costs as % of "static" labour and equipment costs Banks	 Building Societies
1978	0.3	0.9
1979	0.2	2.9
1980	0.4	1.1
1981	1.5	2.1
1982	2.1	12.5
1983	0.6	1.6
1984	0.9	0.3
1985	5.0	4.0
1986	1.0	1.2
1987	1.5	1.7

Table VI: Average Allocative Inefficiencies
Resulting from Adjustment Costs in the UK Banking Sector

Equation (4.22) allows us to calculate the increase in banks' labour and equipment costs that follow from adjustment costs causing banks not to adopt what (in a static model) would be cost minimising behaviour. Note that these cost increases are in addition to the actual adjustment costs themselves and could be considered as allocative inefficiency resulting from the presence of adjustment costs. Table VI displays the results of this calculation averaged over banks and building societies in each year. The table clearly shows the effect of a number of changes to the industry between 1978 and 1987. For the banks these include entry into the mortgage market (1981 and 1982) and the lead up to the *Big Bang* and financial service liberalisation (1985). Meanwhile the building societies progressively built up their presence in the retail banking market in the early 1980s and undertook a number of major mergers in the mid 1980s. The Building Society Act of 1986 allowed them to grow into further areas of financial service activity in 1987.

6. Conclusions

This chapter has set out to achieve a number of objectives. It has taken a dynamic cost function model that incorporates quadratic adjustment costs to changes in the level of factor usages and has applied a small number of plausible assumptions in order to derive an explicit solution to the factor demand equations. Furthermore, it has provided a methodology for estimating the depreciation rate of the capital stock which may differ from that assumed in company accounting procedures.

Both of these additions to the traditional static cost function and factor demand system are necessary when modelling the banking sector. The estimated model is extremely rich in its implications for the UK banking sector and a full discussion of its features in relation to developments in this sector are the subject of the final chapter. In this chapter we have only attempted to outline the features that are particularly pertinent to the theoretical innovations developed in earlier sections - namely the dynamic structure of the cost function and conclusions relating to efficiency.

It is, however, important to compare the principal results of the model estimated in this chapter with previous empirical work. To do this we briefly review our findings with those of the survey of analyses of the US banking sector contained in Clarke (1988) and Ferrier and Lovell (1990). Unfortunately, apart from studies of building societies by Hardwick (1989) and Drake and Weyman-Jones (1992), no comparable studies are available for UK banks. After a review of 13 separate studies Clark concludes that economies of scale exist only for relatively small banks (with less than $100 million in deposits - smaller than any of the banks in our sample). Using two approaches (econometrics and DEA) Ferrier and Lovell find increasing returns to scale for all sizes of bank. The findings in this chapter, that economies of scale exist for retail banks, appear consistent with the findings of Ferrier and Lovell, but initially appear to contradict the results of Clarke. However, the US banking industry is very regionalised (as a result of government regulation), and in this respect may be more comparable to the UK building societies. If banks were allowed to develop in truly national markets they would be able to attain increasing returns to scale.

Our results show that UK building societies do experience decreasing returns to scale once they reach a certain size. This is consistent with the findings of Hardwick (1989) who found increasing returns to scale were exhausted for societies with assets of more than £ 280 million, and significant decreasing returns to scale set in for large societies with assets of more than £1,500 million (such as the ones in our sample). Although Drake and Weyman-Jones (1992) generally accepted constant returns to scale, their DEA did provide evidence that large societies experience decreasing returns to scale. This is again consistent with our findings.

One important conclusion concerning the bank/building society structure of the UK sector is unique to this chapter and worth highlighting.

Analysing data on both banks and building societies has enabled us to identify a notable difference that still exists between the two types of bank. Building societies appear to be managed in such a way as to make them less able to exploit the economies of scale achieved by the retail banks. Presumably we can expect this to change as building societies continue to merge themselves into larger national units and re-structure in order to more effectively compete with retail banks.

Finally, the main objective of this chapter was to develop a theoretical model that could be used to provide a basis against which to measure efficiency improvements in the UK banking sector over time and differences between banks. The model described in this chapter allows for changes in factor prices, adjustment costs and equipment depreciation rates that differ from those used in accounting data. The latter two factors would be particularly difficult to deal with in an approach using numerical calculation of indices. Institutions in the banking sector face particularly high adjustment costs when changing to more equipment intensive methods of service provision and consequent rationalisation of their labour forces. These costs imply that the benefits of more technology intensive methods of service provision are not immediately apparent when employing an essentially static model to measure efficiency improvements. Furthermore, the depreciation rates for equipment used in bank and society accounts over-state the asset lives which banks expect when making investment decisions. This has the effect of over-stating the value of the assets and, therefore, under-stating productivity. The method described in this chapter adjusts for this effect.

Efficiency in both sub-sectors improved only slowly between 1978 and 1979. Between 1979 and 1984 substantial gains were made by both sectors. Since then, efficiency in the two sub-sectors appears to have grown in tandem at a slower rate. The average annual rise in efficiency over the period 1978 to 1987 has been 5.3% for banks and 4.4% for building societies. Both of these estimates are higher than estimates of total factor productivity for the UK economy as a whole over this period.

APPENDIX A: DERIVATION OF EQUILIBRIUM COSTS (A^*_t/A_t)

Pursuing the transformation function of footnote 16, we have

$$h(y_t,z_t,t) = \sum_{n=1}^{N} \beta_n \log(x_{nt})$$

$$= \sum_{n=1}^{N} \beta_n \log(A_t S_{nt}/w_{nt})$$

$$= \log(A_t) \sum_{n=1}^{N} \beta_n + \sum_{n=1}^{N} \beta_n \log(S_{nt}/w_{nt})$$

$$\Rightarrow \quad A_t = \exp\left(h(y_t,z_t,t)/\sum_{n=1}^{N} \beta_n\right) \Big/ \prod_{n=1}^{N} (S_{nt}/w_{nt})^{\beta_n/\sum_{g=1}^{N}\beta_g} \qquad (4.A1)$$

$$\text{Also} \quad A^*_t = \exp\left(h(y_t,z_t,t)/\sum_{n=1}^{N} \beta_n\right) \Big/ \prod_{n=1}^{N} (S^*_{nt}/w_{nt})^{\beta_n/\sum_{g=1}^{N}\beta_g} \qquad (4.A2)$$

And so dividing (4.A1 by 4.A2)

$$(A^*_t/A_t) = \prod_{n=1}^{N} (S_{nt}/S^*_{nt})^{\beta_n/\sum_{g=1}^{N}\beta_g} \qquad (4.A3)$$

Since x^*_t (and, therefore, S^*_{nt} subject to $\sum_{n=1}^{N} S^*_{nt}$) will have been selected to minimise A^*_t, we know that

$$\frac{\partial A^*_t}{\partial S^*_{nt}} \cong -(\beta_n/\sum_{g=1}^{N}\beta_g)A^*_t/S^*_{nt} - \iota = 0 \quad ; \quad n=1,2,\ldots,N \qquad (4.A4)$$

where ι is a Lagrangian Multiplier.

$$\Rightarrow \quad \iota\, S^*_{nt} = -\,(\beta_n / \sum_{g=1}^{N} \beta_g)\, A^*_t \quad ; \quad n=1,2,\ldots,N \tag{4.A5}$$

Summing over all inputs allows us to solve for ι

$$\iota \sum_{n=1}^{N} S^*_{nt} = \iota = -\, A^*_t \tag{4.A6}$$

(since $\sum_{n=1}^{N} S^*_{nt} = 1$)

And so substituting (4.A6) into (4.A5)

$$\beta_n / \sum_{n=1}^{N} \beta_n = S^*_{nt} \quad ; \quad n=1,2,\ldots,N \tag{4.A7}$$

And finally (4.A7) into (4.A3)

$$(A^*_t / A_t) = \pi_{n=1}^{N} \left(S_{nt} / S^*_{nt}\right)^{S^*_{nt}} \tag{4.A8}$$

APPENDIX B: DERIVATION OF USER COST OF CAPITAL

We have that

$$x^+_v - \Delta_1 x^+_{v-1} = x_v - \Delta_2 x_{v-1}$$

Since we can assume $0 \leq \delta_{2n} < 1$ ($n=1,2,\ldots,N$) we can write

$$x_{nv} = \sum_{u=0}^{\infty} \delta^u_{2n} (x^+_{nv-u} - \delta_{1n} x^+_{nv-u-1}) \quad ; \quad n=1,2,\ldots,N$$

$$= x^+_{nv} + \sum_{u=1}^{\infty} (\delta^u_{2n} - \delta_{1n}\delta^{u-1}_{2n})\, x^+_{nv-u}$$

We now wish to find a price associated with a unit of factor service x^+_{nv}. Let this price be w^+_{nv}. Then the discounted future factor payments are

$$\sum_{v=t}^{\infty} \sum_{n=1}^{N} w_{nv} x_{nv}/(1+r)^{v-t}$$
$$= \sum_{v=t}^{\infty} \sum_{n=1}^{N} w_{nv}\Big(x^+_{nv} + \sum_{u=1}^{\infty} (\delta^u_{2n} - \delta_{1n}\delta^{u-1}_{2n}) x^+_{nv-u}\Big)/(1+r)^{v-t}$$

Re-arranging to collect all terms in x^+_{nv} ($n=1,\ldots,N$; $v=t,\ldots,\infty$) on the right hand side

$$\sum_{v=t}^{\infty} \sum_{n=1}^{N} w_{nv} x_{nv}/(1+r)^{v-t}$$
$$- \sum_{v=t}^{\infty} \sum_{n=1}^{N} w_{nv}\Big(\sum_{u=v-t+1}^{\infty} (\delta^u_{2n} - \delta_{1n}\delta^{u-1}_{2n}) x^+_{nv-u}\Big)/(1+r)^{v-t}$$

$$= \sum_{v=t}^{\infty} \sum_{n=1}^{N} w_{nv}\Big(x^+_{nv} + \sum_{u=1}^{v-t} (\delta^u_{2n} - \delta_{1n}\delta^{u-1}_{2n}) x^+_{nv-u}\Big)/(1+r)^{v-t}$$

$$= \sum_{v=t}^{\infty} \sum_{n=1}^{N} \Big(w_{nv}/(1+r)^{v-t} + \sum_{u=1}^{\infty} (\delta^u_{2n} - \delta_{1n}\delta^{u-1}_{2n}) w_{nv+u}/(1+r)^{v-t+u}\Big) x^+_{nv}$$

$$\equiv \sum_{v=t}^{\infty} \sum_{n=1}^{N} w^{+}_{nv} x^{+}_{nv}$$

since $\sum_{v=t}^{\infty} \sum_{n=1}^{N} w_{nv}\Big(\sum_{u=v-t+1}^{\infty} (\delta_{2n}^{u} - \delta_{1n}\delta_{2n}^{u-1}) x^{+}_{nv-u}\Big)/(1+r)^{v-t}$ contains no terms in x^{+}_{nv+u} for $u>0$.

$$\therefore \quad w^{+}_{nv} = w_{nv}/(1+r)^{v-t} + \sum_{u=1}^{\infty} (\delta_{2n}^{u} - \delta_{1n}\delta_{2n}^{u-1}) w_{nv+u}/(1+r)^{v-t+u}$$

In particular when $v=t$

$$w_{nt} = w_{nt} + \sum_{u=1}^{\infty} (\delta_{2n}^{u} - \delta_{1n}\delta_{2n}^{u-1}) w_{nt+u}/(1+r)^{u}$$

$$= w_{nt} + (1-\delta_{1n}/\delta_{2n}) \sum_{u=1}^{\infty} \big(\delta_{2n}/(1+r)\big)^{u} w_{nt+u}$$

Replacing future prices by their expectations (from equation (4.12a))

$$w^{+}_{nt} = w_{nt} \Big(1 + (1-\delta_{1n}/\delta_{2n}) \sum_{u=1}^{\infty} (\delta_{2n}\zeta_{1n})^{u} \Big)$$

$$= w_{nt} \Big(1 + \zeta_{1n}(\delta_{2n}-\delta_{1n})/(1-\delta_{2n}\zeta_{1n}) \Big) \quad \text{(assuming } 0 \leq \delta_{2n}\zeta_{1n} < 1\text{)}$$

$$= w_{nt} \; (1-\delta_{1n}\zeta_{1n})/(1-\delta_{2n}\zeta_{1n})$$

$$= \frac{1-\delta_{1n}\zeta_{1n}}{1-\delta_{2n}\zeta_{1n}} \, w_{nt}$$

Chapter V

Summary and Conclusions

ABSTRACT

We compare direct cost efficiency indices and parametric models as a means of estimating cost efficiency when there are adjustment costs. We prefer a parametric model based approach. In the case of the application of the methods to the banking and building society sectors the importance of adjustment costs is discussed.

The conclusions of this thesis relate to two separate themes. First, we present methodological comparisons and conclusions relating to two approaches to cost efficiency measurement in the context of a dynamic model of the firm. Second, we present the main conclusions from the empirical application of these two methodologies to the UK banking sector between the years 1978 and 1987, particularly relating to the impact of changes in the scale and methods of service provision on cost efficiency performance.

1. Two Approaches to Efficiency Measurement

This thesis has majored on estimating the differences in cost efficiency between firms, or between time periods, when there are adjustment costs associated with a changed scale of production from the previous time period (in terms of changes in factor inputs). These adjustment costs can be an important component of the cost structure of firms in some industries. More specifically, we have shown them to be important in the bank and building society sector.

The principal challenge for any approach to cost efficiency comparisons is to find a method that adjusts the cost base of a firm for a number of factors that are essentially beyond the control of the firm and that may differ, between time periods or firms, from a specified reference. These may include the level of output, factor input prices and, in our context, the previous time period's factor inputs (which will determine current period adjustment costs) and expectations of future factor input prices and demands.

Our approach in this thesis has recognised that in a situation where there are adjustment costs, we need to specify inefficiency so that unavoidable adjustment costs resulting from a change in the firm's exogenous output or factor input prices are excluded. That is, we have excluded adjustment costs that are incurred by the firm when it is minimising total costs (factor payments and adjustment costs). Therefore, we have defined inefficiency to result only from those costs that are in excess of this minimum level.

In chapters III and IV we have made a detailed analysis of two approaches to efficiency measurement. First, we have considered adaptations of the indices first analysed in detail by Caves, Christensen and Diewert (1982). Second, we have considered parametric approaches using an econometric model of the firm's cost structure. In fact both are based on an underlying parametric model of the firm or firms being compared. However, we have shown that the index number approach avoids the need to explicitly estimate the static components of this model, whilst still allowing it to take a flexible form that is hoped will approximate the production technology and cost structure for both the firms under study and the reference against which it is being compared. However, the disadvantage of this approach is that it does not give an understanding of the structure of the adjustment cost process, and in particular does not allow us to analyse the role of expectations in the firms' decisions determining current period factor demands. By contrast, the parametric approach we have developed explicitly estimates a structural model suitable for a reference cost structure, including adjustment costs and even the firms' expectations forming process.

Although both approaches have been well developed for static models of the firm, new adaptations have been necessary for the case considered in this thesis where the cost structure includes adjustment costs. We will now summarise both approaches, and the adaptations to deal with adjustment costs.

In the case of the index number approach we have found that, by assuming myopic behaviour for the firm, a modified index can be derived that is essentially the usual static index augmented by some additional terms that relate to the adjustment costs and their derivatives. We have shown that if both the specification of the adjustment costs and the value of all the parameters within that specification are known, then an adjusted cost efficiency index can be derived. The disadvantage of this is that in virtually all practical situations this information will not be available. Furthermore, use of a rigid specification (including parameter values) for the adjustment costs is contrary to the original motivation for the index number approach (namely to reduce dependency on parametric formulations to as general and flexible a model as possible). However, we have argued that it is possible to substitute the unknown derivatives of the adjustment costs that are included in the augmented index for a particular parametric reduced form. This can then be estimated as an *econometric model* in which the usual static cost efficiency index (incorporating current period costs, output and factor prices) is decomposed into two components. The first is a deterministic component, expressing *static inefficiency* resulting from adjustment costs (identified by a reduced form flexible function). The second is a model residual, capturing *other* inefficiencies. Effectively, we are using a static cost efficiency index to estimate both inefficiencies and adjustment costs together, and then using a reduced form econometric model to separate out the two. The resulting estimates of inefficiency are short term measures, conditional on the last periods' factor usages which are taken as exogenous.

However, there is a pragmatic drawback to this quasi-index number approach. Since the nature of adjustment costs in the sector being studied may be unknown, it is important that there exists some procedure for verifying that any assumed specification is indeed appropriate. Furthermore, in most applied work it will be extremely helpful to quantify the impact of adjustment costs on the firms being studied.

In the parametric approach adopted in chapter IV, we have shown how a specification for adjustment costs can be incorporated in the structural model of a firm's cost function and explicitly estimated. We have also shown how an expectations forming process can be specified whereby the firm takes account of expected future trends in its factor demands in deciding upon current levels (minimising the present value of current and expected future costs). This allows us to estimate the long term cost structure of the firms in a sample, to be used as a reference against which efficiency is measured. We have argued that it is possible to estimate the size and test the significance of the structural parameters associated with adjustments of individual factor inputs (or outputs[21]). Furthermore, we have shown that it is possible to separately estimate the overall contribution of adjustment costs to the firms' total costs, and the amount by which total costs rise as a consequence of adjustment costs. (This may be more than the actual amount of adjustment costs since firms may have opted for higher factor payments in order to benefit from lower adjustment costs).

The different approaches of chapters III and IV also led us to distinguish between short and long term inefficiency. In chapter III

---

[21] Although adjustment costs associated with changes in outputs were not included in the models in chapters III or IV, this extension would be very easy. This is because both the current and previous period output levels (y_t and y_{t-1}) are assumed to be exogenous variables. Therefore, any adjustment cost function of y_t and y_{t-1} would also be exogenous, outside the firm's control.

we took as given the firm's factor usage in the previous production period and, therefore, only looked at the firm's performance in the current period. However, in chapter IV we took a long term view of inefficiency by measuring how far away the firm was from the minimum cost it could have achieved if it behaved efficiently throughout its history.

Of course, the two approaches should give consistent (although not identical) estimates of overall cost efficiency. We would expect differences since one approach estimates short run efficiency (the index number approach), whilst the other estimates long run efficiency (the parametric model approach). Furthermore, the explicit functional forms chosen to model the underlying cost function and adjustment costs in the parametric approach may differ from the implicit forms assumed in the direct index number approach.

The empirical work carried out in this thesis did produce similar results for the two approaches in the sample of UK building societies (in fact the rankings of cost efficiency differed in only one instance, reversing second and third place). However, there were some striking differences for the sample of retail banks. The reason for this appears to lie in the greater range of outputs that were included in the structural econometric approach. This is due to the fact that in the index number analysis of chapter III, only one output (real value of accounts) was considered, since a greater number would have involved an excessive number of terms with consequent multicollinearity problems in the estimation. However, in the structural parametric approach (of chapter IV), a total of three outputs were successfully incorporated into the model. This was possible partly because of the additional parameter restrictions that could be imposed as a result of using a structural model (in particular, adding-up of factor shares and homogeneity of prices). The greater number of outputs in the structural econometric approach alters our rankings of the cost efficiency of a number of banks which produce diverse output combinations. This was not so important in the case of building societies which have more homogeneous businesses, maintaining approximately the same balance between accounts and advances (largely a necessity of building societies being denied access to wholesale funds during our modelling period). This is shown by the graphs of outputs for banks and building societies in Annex 1 of this thesis (Charts A.I and A.II).

2. Empirical Findings in the Banking Sector

The empirical work in both chapters III and IV (using different methods, and to some extent different models) identified significant cost efficiency improvements for both banks and building societies over the period from 1978 to 1987. In the case of the preferred parametric approach in chapter IV, these efficiency improvements averaged 5.3% per annum for banks and 4.4% per annum for building societies.

These gains in efficiency have been estimated after taking account of adjustment costs that have occurred in this fast growing sector of the economy. Incorporating adjustment costs into the structural parametric model of chapter IV has enabled us to estimate the importance of these dynamic effects for the UK bank and building society sector. Surprisingly, we found that adjustments to the size of the labour force used by a bank or building society were more important than adjustments to its stock of plant and equipment. (A 10% change in the stock of labour resulted in an adjustment cost of 6.5% on top of the bank or building society's usual wage bill, whilst, in the case of plant and equipment, the additional cost was only 1.9% on top of the user cost for plant and equipment.) Clearly, changes in the size of the labour force directly resulted in some easily identifiable costs (most obviously recruitment, training and redundancy payments - which in most cases in the bank and building sector are large). However, it also seems possible that there were indirect costs associated with the substitution between labour and computers. Reductions in the size of the labour force made possible by the installation of new computer systems may have resulted in redundancy payments, but also required re-training of the remaining staff so that they may operate the new systems. In this respect separability of adjustment costs into those resulting from changes in labour and computers may not be meaningful.

The importance of adjustment costs for the banking sector has been estimated in chapter IV. For the industry as a whole over the period from 1978 to 1987, adjustment costs added between 0.9% and 2.9% to labour costs and between 1.4% and 6.9% to equipment costs. The level of adjustment costs varied considerably from year to year depending on general growth and structural changes within the industry. For example, *Big Bang*, the entry of banks into the mortgage lending market and the loosening of restrictions on the activities of building societies during the 1980s all had discernible impacts on the adjustment costs faced by the respective bank types within the sector.

An efficient bank should lessen the impact of adjustment costs by re-distributing factor usages towards those with lower adjustment costs, so as to minimise both factor payments *and* adjustment costs. It therefore follows that "raw" adjustment costs alone will under-estimate the final impact on costs - since we need to add in the cost of additional factor payments as banks re-arrange their inputs from "static optimal" levels. In chapter IV we estimated that these additional factor payments add between zero and 5% to bank costs and between zero and $12^1/_2$% to building society costs.

Finally, the results of this thesis suggest three explanations for the absence in previous empirical work of any identifiable short-term productivity benefit from investment in information technology equipment. First, the conclusion from the empirical analysis in chapter IV, that adjustment costs for labour are higher than those for plant and equipment, means that efficient banks and building societies attempt to meet a rising output (as shown by 87.7% real growth for the sector as a whole between 1978 to 1987 period) through a relatively greater investment in plant and equipment. This results in an increase in the ratio of plant and equipment per unit of output. Second, the extra investment in plant and equipment still incurs a significant adjustment cost for the bank or building society in the short term. Third, the model estimated in chapter IV has shown there to be significant evidence that general bank and building society sector accounting assumptions of five year lives for plant and equipment will have over-estimated the actual useful lives of the assets. On the basis of our data from 1978 to 1987, a more appropriate figure would have been 4 years or less. This in turn will have resulted in an over-estimate of the value of the stock of plant and equipment held by banks and building societies. These three factors taken together may explain the absence in previous empirical work of any identifiable short-term productivity benefit from investment in information technology equipment.

ANNEX 1: SOURCES AND DESCRIPTION OF THE DATA

In this annex we describe the sources and data used for outputs and factor inputs of banks and building societies in the modelling work of chapters III and IV.

(a) Outputs

We adopt a production approach and attempt to find variables that will proxy the unobserved level of transactions that constitute the service of deposit and loan accounts. In the case of a bank or building society, we may consider output to be generated by transactions associated with:

- receiving deposits;
- repaying deposits;
- making loans;
- recovering loans;
- paying interest;
- receiving interest;
- managing accounts;
- managing loans.

All of these activities involve various degrees of complexity.

Financial data indicating the level of activity is available from company and society annual reports. Such data includes:

- interest paid on accounts;
- interest received on loans;
- value of accounts;
- value of advances;
- value of leased assets.

Data on the value of accounts, interest payments, and advances was deflated by the Retail Prices Index to give the three output measures covering the most important areas of activity. This approach is extremely appealing due to its richness in allowing the model to capture the effect of the different output mixes that the banks provide.

Charts A.I and A.II display aspects of the output data. From these charts it will be noted that the real levels of income received and interest payed are roughly proportional to the value of accounts, although some variation is evident. This variation is larger in the case of banks than of building societies, indicating that building societies are more homogeneous in their output mixes.

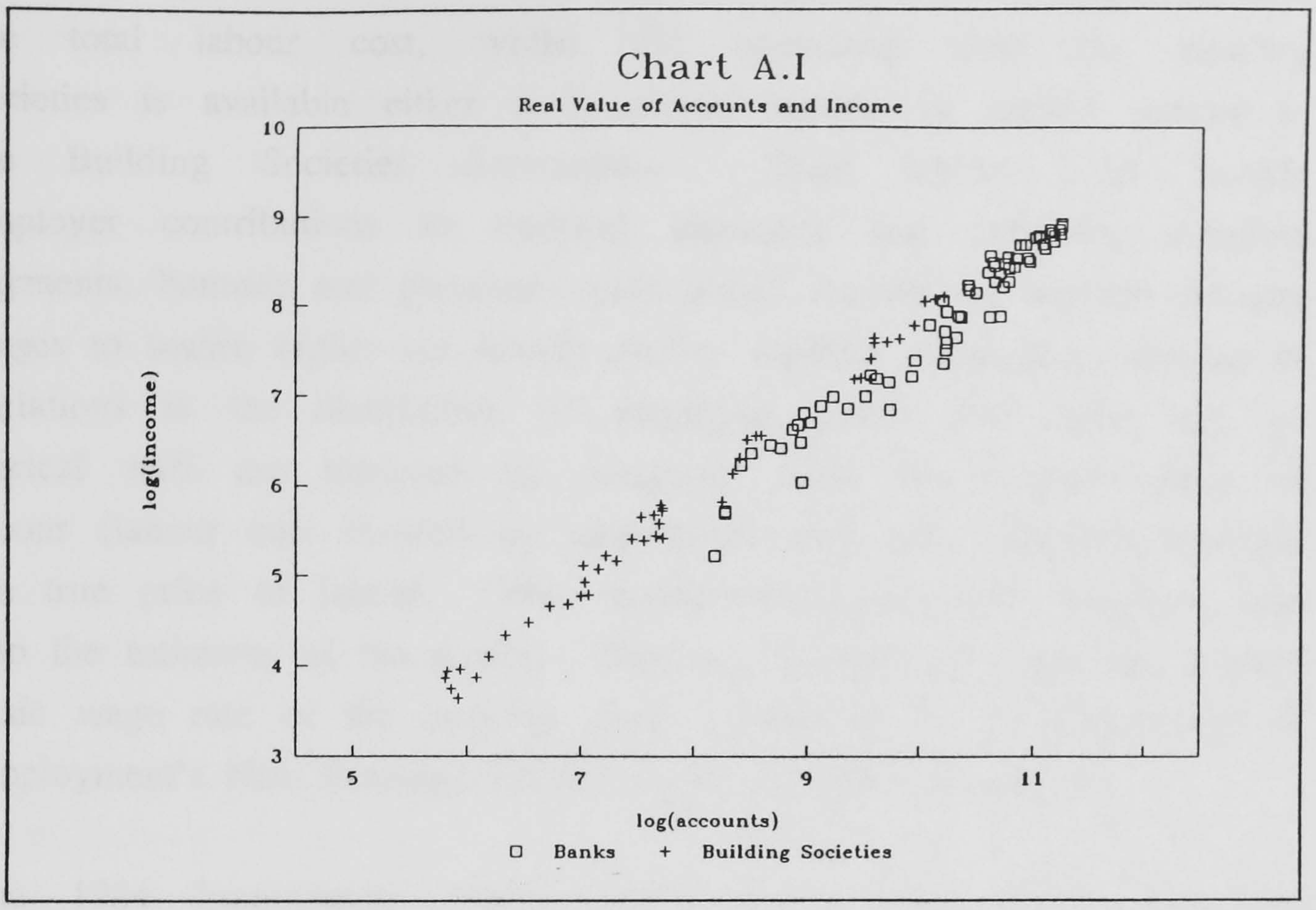


Chart A.I
Real Value of Accounts and Income
log(income)
log(accounts)
Banks
Building Societies

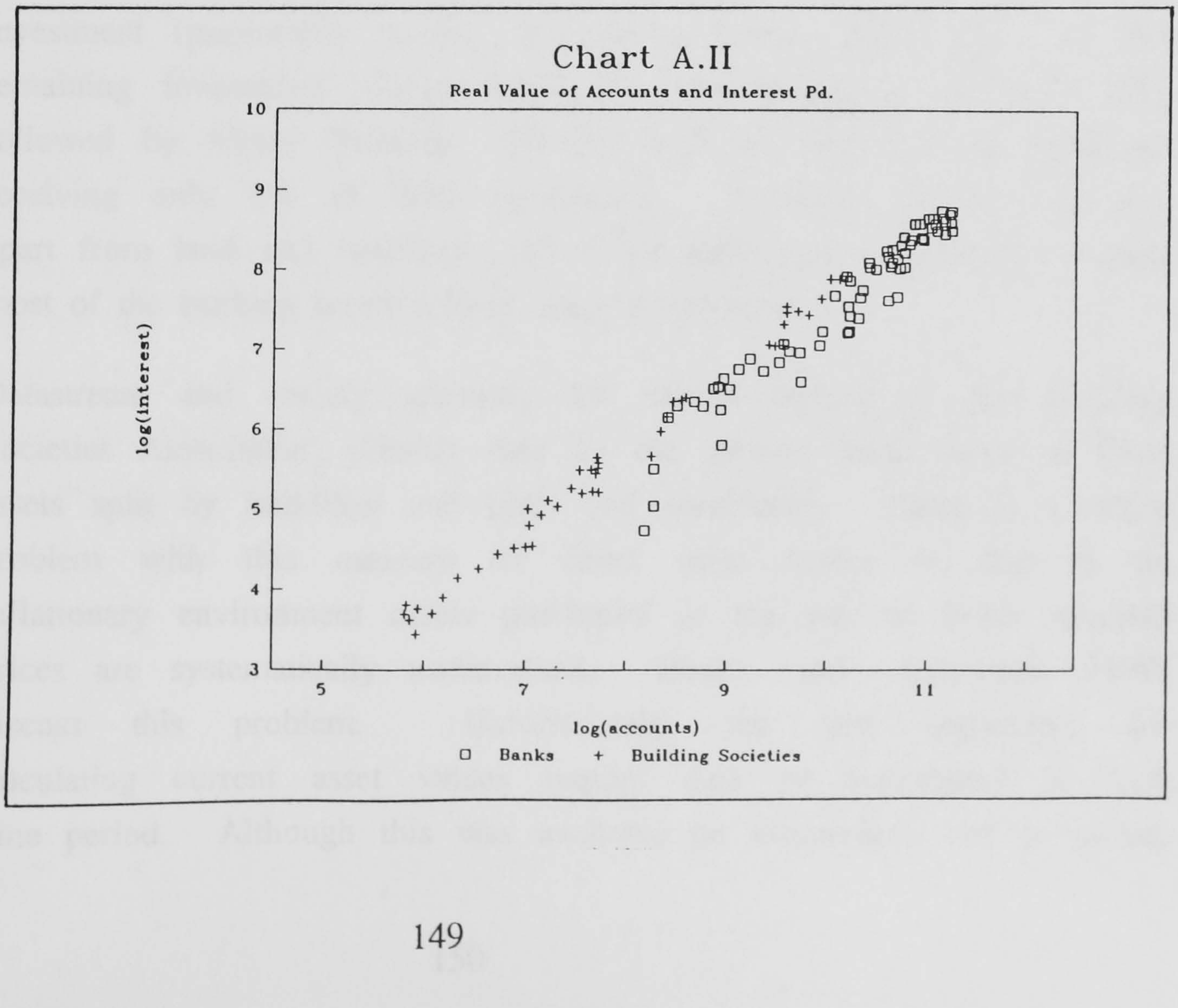


Chart A.II
Real Value of Accounts and Interest Pd.
log(interest)
log(accounts)
Banks
Building Societies

(b) Inputs

Datastream provides data for each bank on the number of employees and the total labour cost, whilst the equivalent data for building societies is available either from annual reports or annual returns to the Building Societies Association. Total labour costs include employer contributions to national insurance and pensions, overtime payments, bonuses and premiums paid above (or below) average industry wages to secure higher (or lower) quality workers. However, because of variations in the distribution of employee grades over time (e.g. as clerical staff are replaced by computer staff) the implied price of labour (labour cost divided by employees) may not accurately measure the true price of labour. This measurement error will introduce bias into the estimates of the model. This was avoided by using the average basic wage rate of the banking sector (measured in the Department of Employment's New Earnings Survey) as an instrumental variable.

The 1984 Input/Output Tables published by the Central Statistical Office purport to give Gross Domestic Fixed Capital Formation for the banking sector analysed by the sector from which the capital was purchased. This shows that the largest beneficiary of investment by the banking sector is the construction sector, receiving 60% of total investment (presumably mostly to enhance branch networks). Of the remaining investment, Office Machinery and Computers receives 25%, followed by Motor Vehicles with 5% and all other sectors combined receiving only 4% of total investment. Therefore, it is clear that apart from land and buildings, office machinery and computers comprise most of the banking sector's fixed asset investments.

Datastream and society accounts (or annual returns to the Building Societies Association) provide data on the historic book value of fixed assets split by buildings and plant and machinery. There is a major problem with this measure of fixed asset stocks in that in an inflationary environment assets purchased in the past at lower nominal prices are systematically undervalued. Bond and Devereux (1990) discuss this problem. Unfortunately, the best algorithms for calculating current asset values require data on investment in each time period. Although this was available on Datastream and in society accounts for all assets in total, it was not disaggregated to individual classes of assets. Investments in each asset class do not always follow the same trends. For our purposes it was particularly important to distinguish buildings from plant and machinery, since the later reflected information technology with very different characteristics to buildings (e.g. in terms of asset life and substitutability for labour). However, knowledge of investment in each period is not necessary if we are prepared to make the assumption that assets are purchased at a steady rate. Then, the volume of assets employed may be estimated by historic book values deflated by an average of a relevant price index, averaged over the assets' lives. The relevant price index in the case of plant and machinery was taken to be the Electrical and Electronic Engineering Producer Output Price Index published by the Department of Trade and Industry.

Rented plant and machinery also provide services to most banks and need to be included in the asset stock. This was done by assuming that the rental payments equate to the user cost (described below), so that the implied value of the asset could be computed and added onto the purchased asset stock.

Data on the number of branches operated by banks and building societies is available from company reports and accounts. This was thought to be a better measure of the volume of buildings used for operational purposes than the value of land and buildings since other buildings may be held purely as investments.

Charts A.III, A.IV and A.V display the input data in comparison with one of the output indicators (the real value of accounts). It will be noted that the size of branch networks appear rigid over time for most banks and building societies. Likewise, numbers of employees have a tendency not to increase over time in line with output.

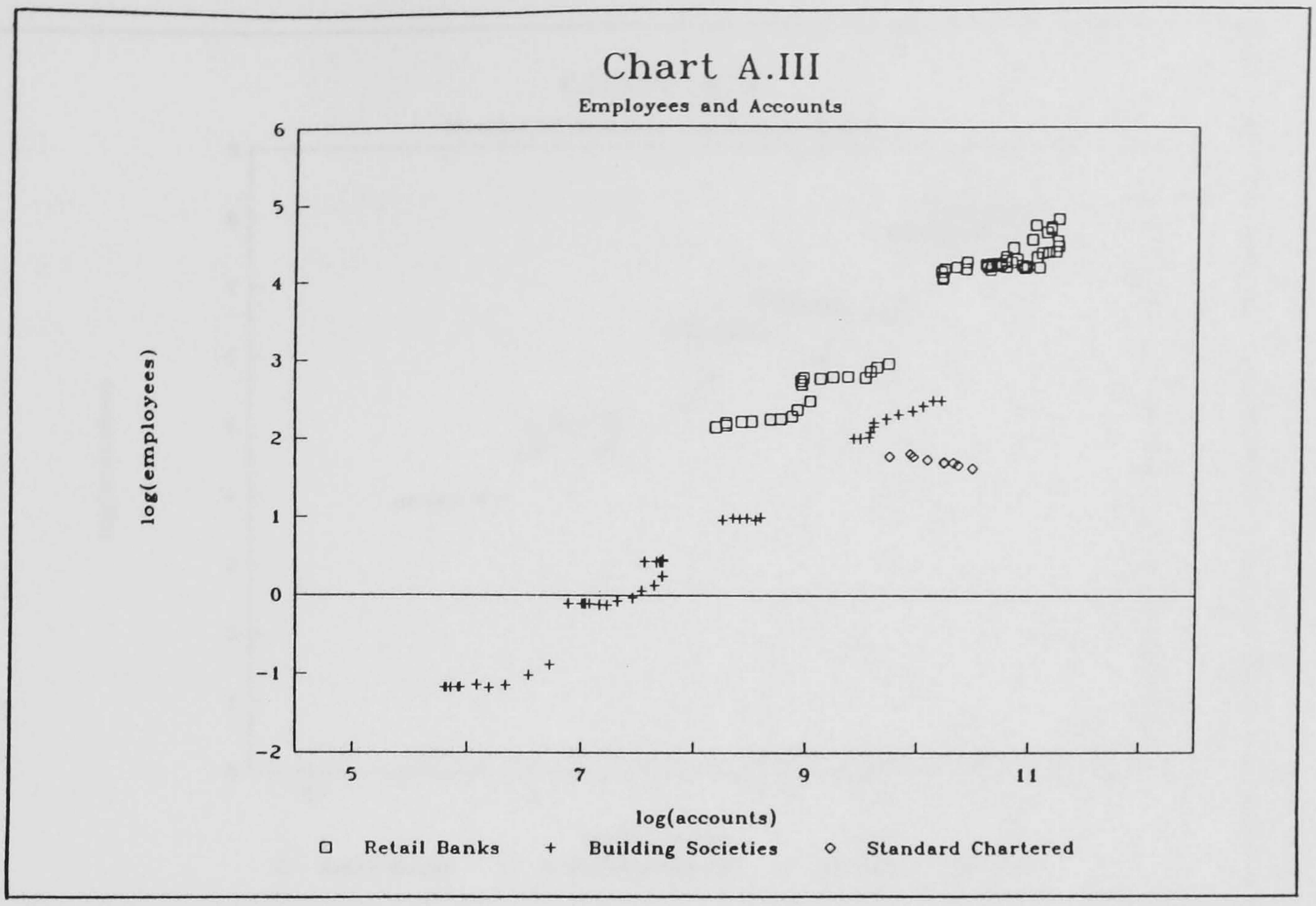


Chart A.III
Employees and Accounts
log(employees)
log(accounts)
Retail Banks
Building Societies
Standard Chartered

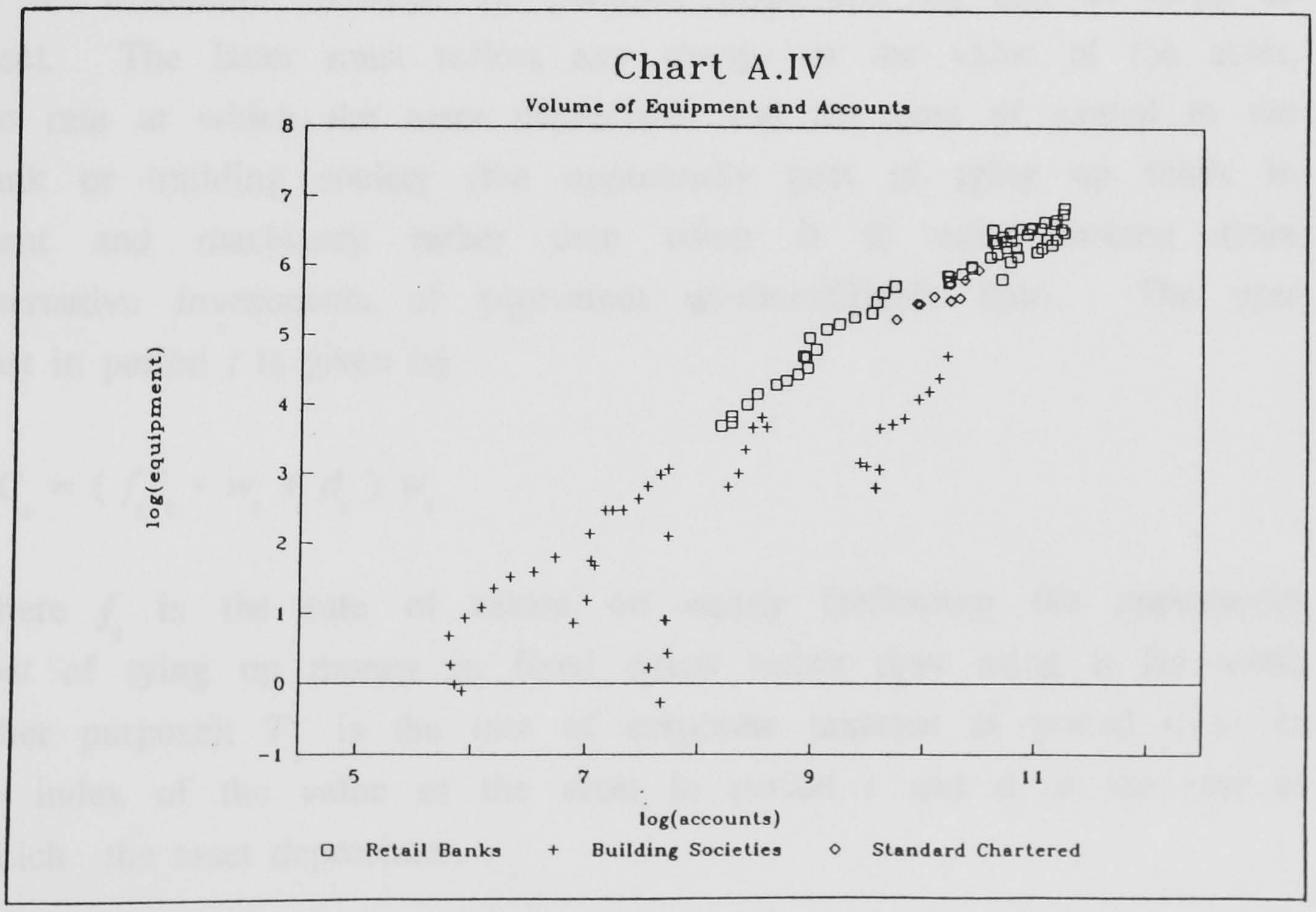
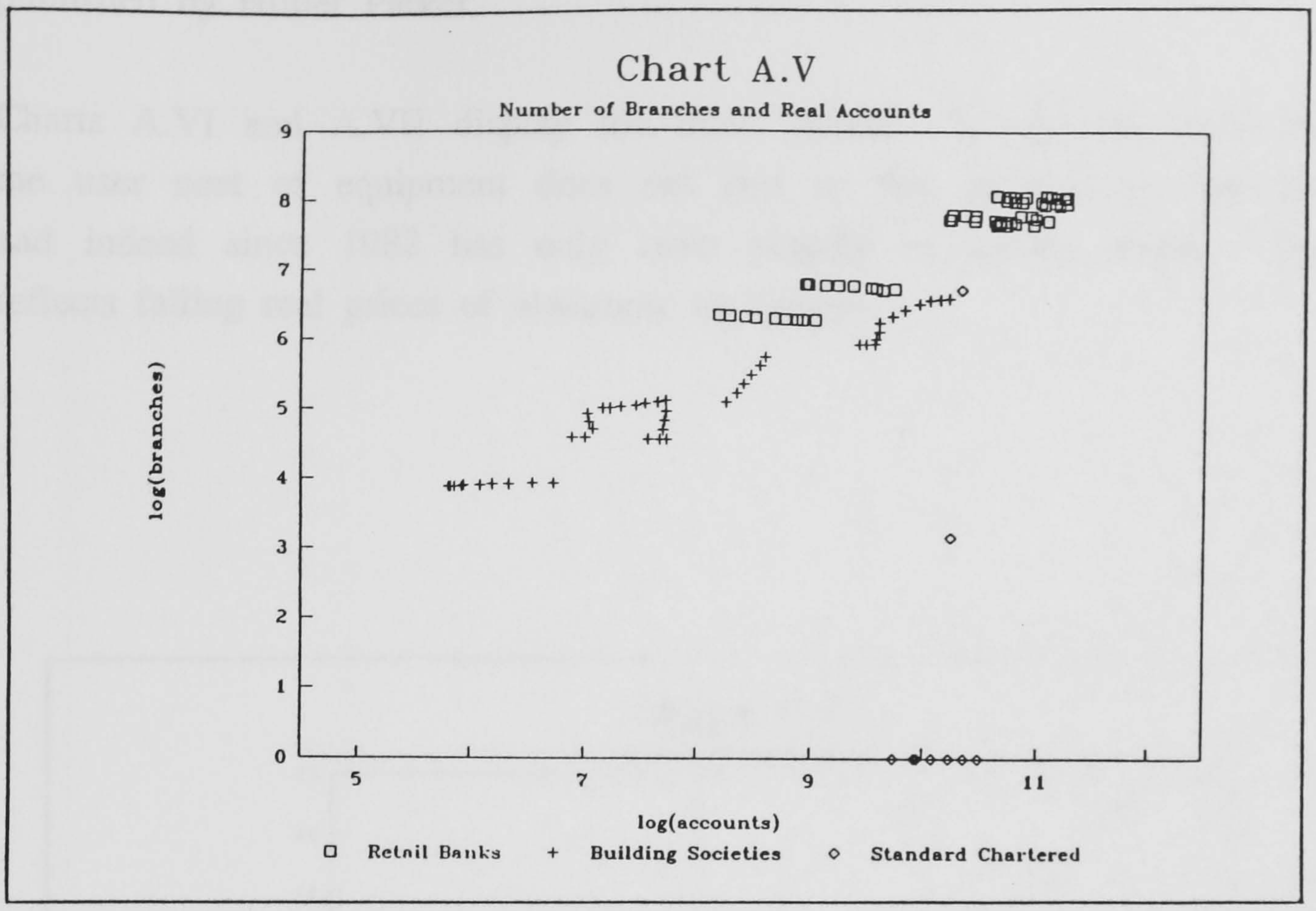


Chart A.IV
Volume of Equipment and Accounts
log(equipment)
log(accounts)
Retail Banks
Building Societies
Standard Chartered



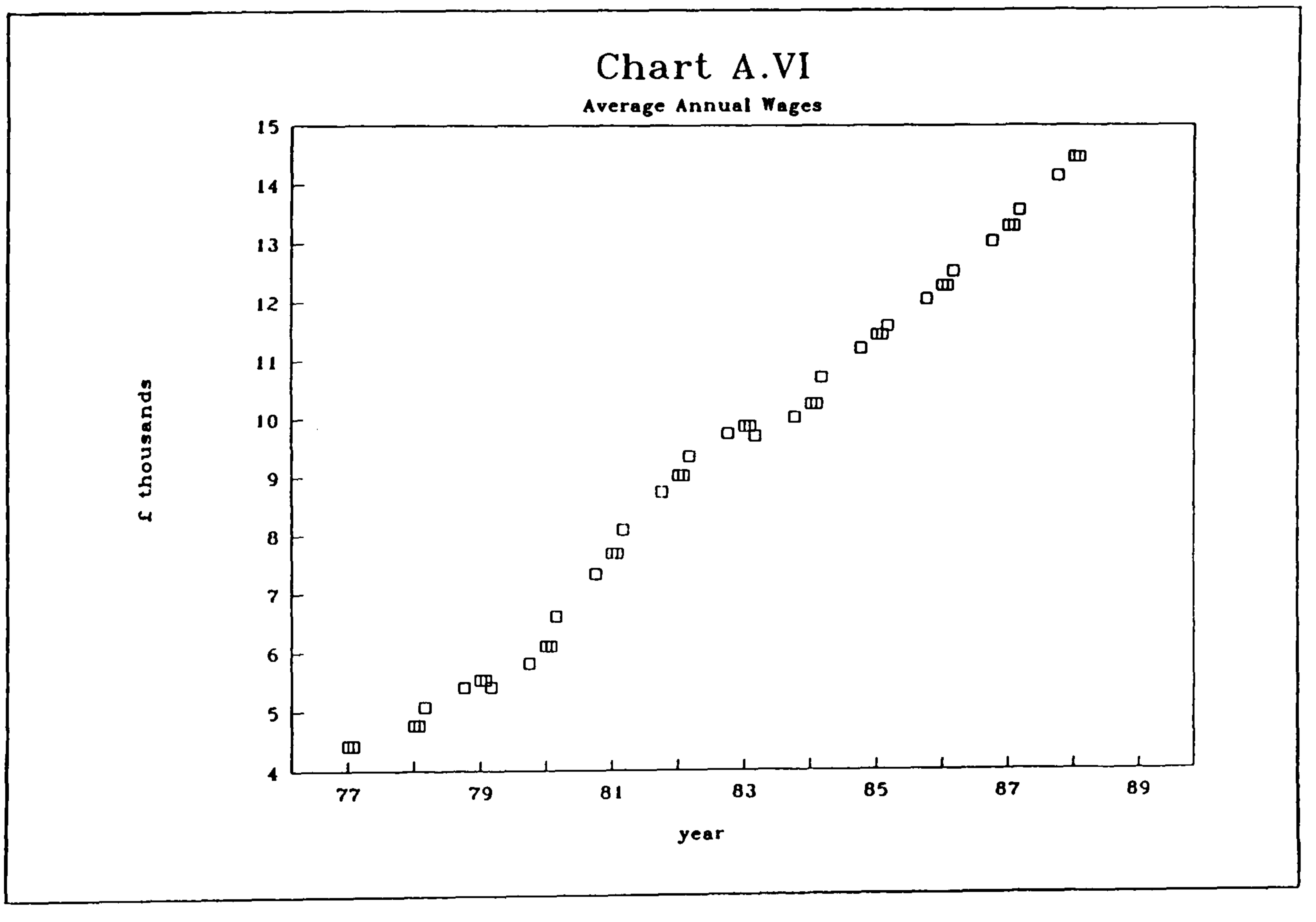
The price of plant and machinery required careful consideration because of the distinction between the purchase price and the cost of using an asset. The latter must reflect any change in the value of the asset, the rate at which the asset depreciates and the cost of capital to the bank or building society (the opportunity cost of tying up funds in plant and machinery rather than using it to raise interest from alternative investments of equivalent un-diversifiable risk). The user cost in period t is given by

$$UC_t = (f_t T_t - w_t + d_t) \, w_t$$

where f_t is the rate of return on equity (reflecting the opportunity cost of tying up money in fixed assets rather than using it for some other purpose); T_t is the rate of corporate taxation in period t; w_t is an index of the value of the asset in period t and d_t is the rate at which the asset depreciates.

The user cost for branches was simply taken to be the office rent index published by Hillier Parker.

Charts A.VI and A.VII display the input prices. It will be noted that the user cost of equipment does not rise as fast as that of branches, and indeed since 1982 has only risen slightly in money terms. This reflects falling real prices of electronic equipment.



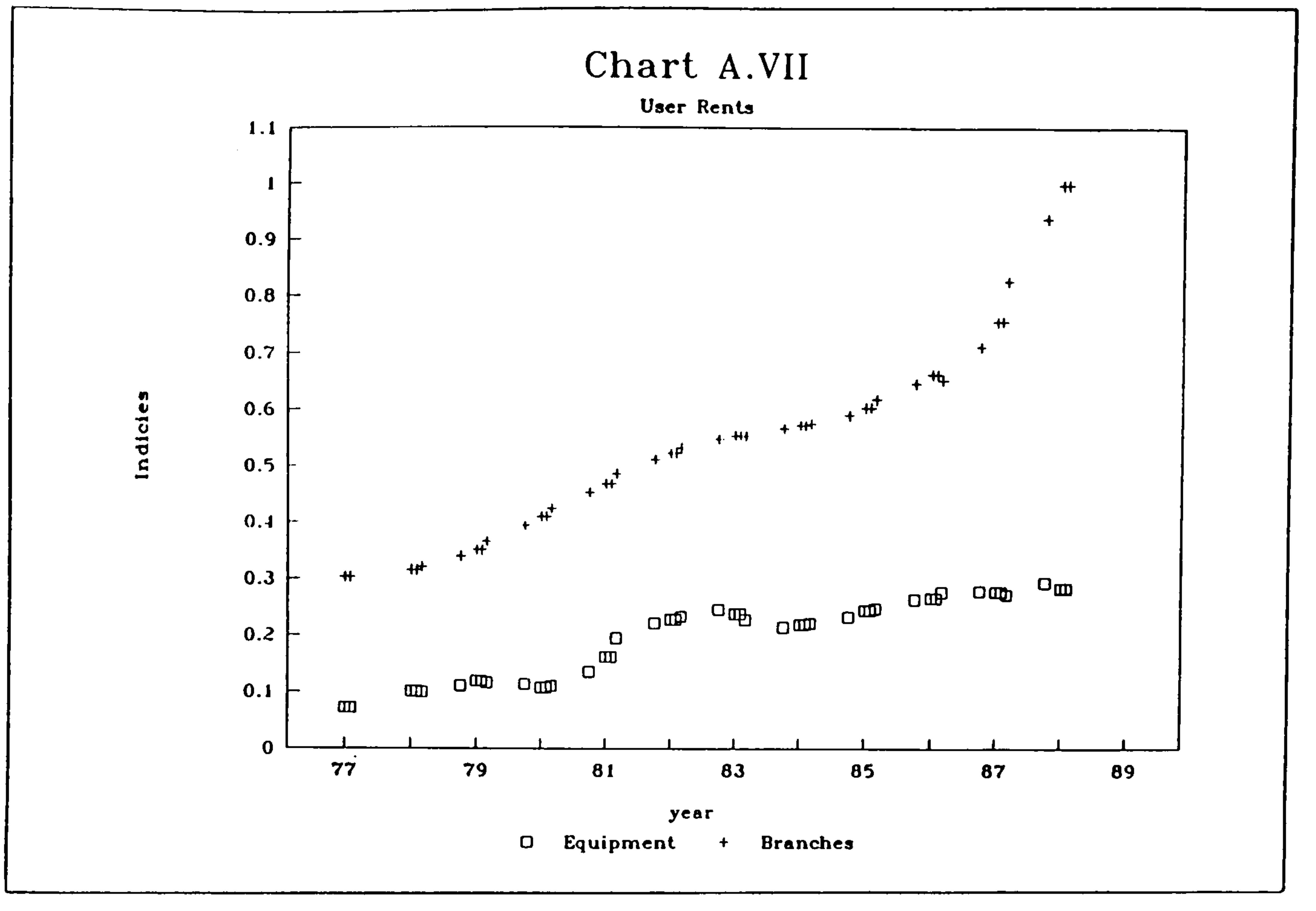
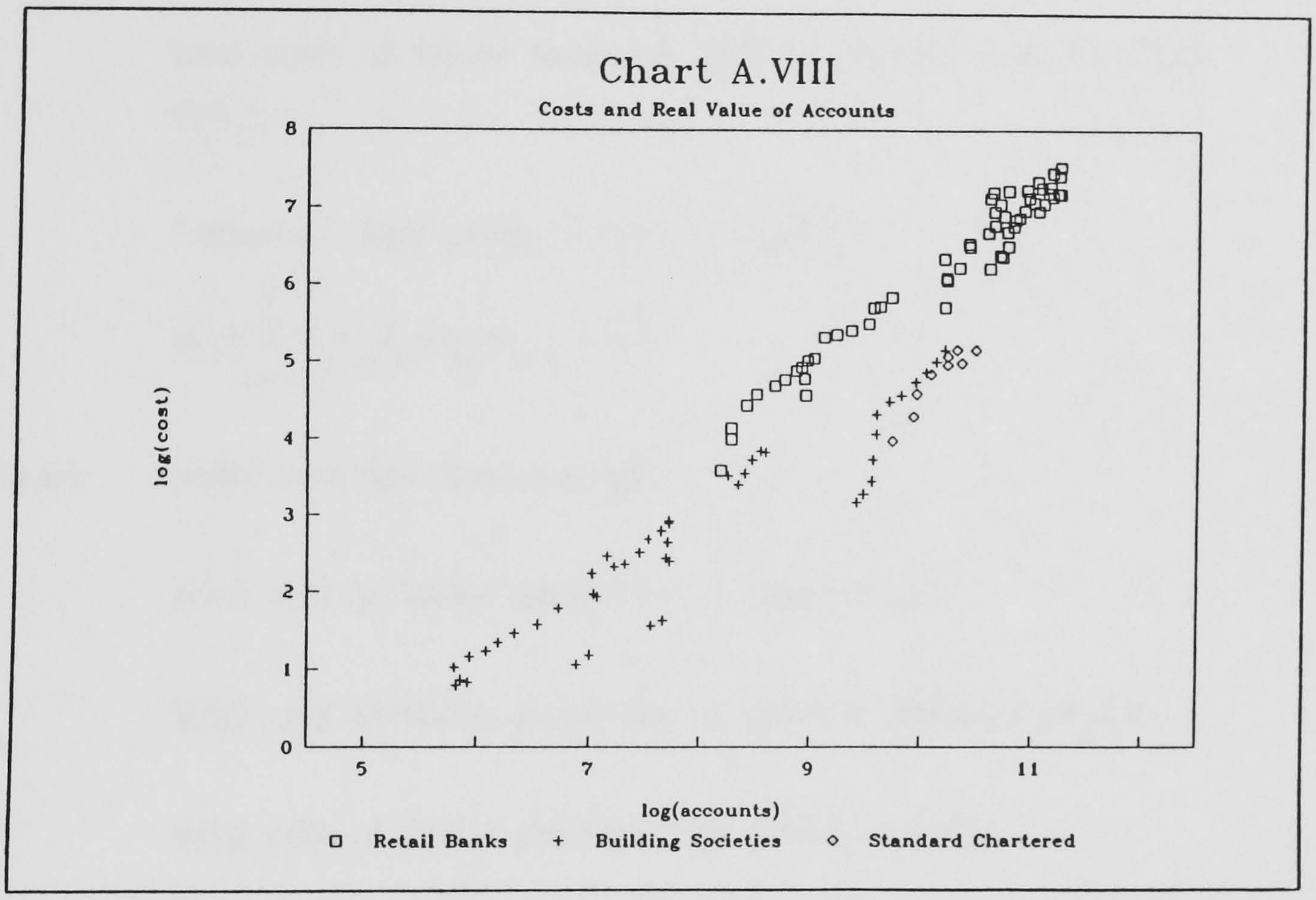


Chart A.VII
User Rents
Indicies
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
77
79
81
83
85
87
89
year
Equipment
Branches

Chart A.VIII displays the total costs of labour, and plant and machinery to the banks and building societies in the sample, against one of the output indicators (the real value of accounts). A clear relationship can be seen.



ANNEX 2: GLOSSARY OF PRINCIPAL NOTATION

$a^s(y_t,w_t,x_t)$ cost function for firm s incorporating adjustment costs;

a_t total cost of factor payments and adjustment costs for firm in period t only;

A_t total costs of factor payments and adjustment costs for firm $(\gamma_t C_t)$;

B_t "adjusted" total costs of firm in period t $(C_t + \sum_{n=1}^{N} \pi_n w_{nt} x_{nt} \{x_{nt}/x_{nt-1} - 1 - \zeta_n\})$;

$c^s(y,w)$ static cost function of firm s;

C total cost of factor payments for firm $(w.x)$;

C_t total cost of factor payments for firm in period t $(w_t.x_t)$;

C^s total cost of factor payments for firm s $(w^s.x^s)$;

C^s_t total cost of factor payments for firm s in period t $(w^s_t.x^s_t)$;

$\breve{C}$ mean of total factor payments (C^s_t) averaged across all firms in sample;

$\tilde{C}^s_t$ "normalised" total factor payments for firm s in period t $(C^s_t/\breve{C})$;

$d^s(x,y)$ distance function used in defining technical efficiency of firm s in producing output y from factor inputs x; function evaluates to a scalar;

d_{1s}	dummy variable taking on value of one for observations on firm s and zero otherwise;
d_{2t}	dummy variable taking on value of one for observations in period t and zero otherwise;
$e^s(c,y,w)$	distance function used in defining cost efficiency of firm s in producing output y at cost C when factor input prices are w; function evaluates to a scalar;
$f^s(x,y)=0$	transformation function for firm s to produce output y from factor input x;
g	alternative subscript reference for factor inputs used by firm;
$G(.)$	variable cost function inclusive of adjustment costs, as used by Morrison and Berndt (1981 and 1991);
h	alternative subscript reference for outputs produced by firm;
i	alternative subscript reference for factor inputs used by firm;
j	alternative subscript reference for outputs produced by firm;
m	subscript reference for outputs produced by firm;
M	number of outputs produced by firm;
n	subscript reference for factor inputs used by firm;
N	number of factor inputs used by firm;
p	vector of output prices of firm $(p_1,p_2,\ldots,p_M)$;

p_t vector of output prices of firm in period t $(p_{1t},p_{2t},...,p_{Mt})$;

p^s vector of output prices of firm s $(p^s_1,p^s_2,...,p^s_M)$;

p^s_t vector of output prices of firm s in period t $(p^s_{1t},p^s_{2t},...,p^s_{Mt})$;

$P^s(x,y)$ technical efficiency of producing output y from factor input x, measured against a reference technology s;

$P^s(x^k,x^l,y^k,y^l)$ index for technical efficiency of producing output y^l from factor input x^l compared to producing y^k from x^k, based on reference technology of firm s;

Q_{st} estimated cost efficiency of firm s in period t;

$Q^s(C,y,w)$ cost efficiency of producing output y at cost C when factor input prices are w, measured against a reference technology s;

$Q^s(C^k,C^l,y^k,y^l,w^k,w^l)$ index for cost efficiency of producing output y^l at cost C^l when factor input prices are w^l compared to producing y^k at cost C^k when factor input prices are w^k, based on reference technology of firm s;

r discount rate;

$R^s(x,y,C,w)$ allocative efficiency of using factor input x to produce output y at cost C when factor input prices are w, measured against a reference technology s;

S^s_n share of firm s total cost incurred from factor input n $(w^s_n x^s_n / w^s . x^s)$:
S^s_l is firm s factor share for labour;
S^s_k is firm s factor share for equipment;
S^s_b is firm s factor share for branches;

S^s_{nt} share of firm s total cost incurred from factor input n in period t $(w^s_{nt} x^s_{nt} / w^s_t . x^s_t)$;

$\bar{S}_n$ share of total cost incurred from factor input n for average of all firms in sample $(\bar{w}^s_{nt} \bar{x}^s_{nt} / \bar{w}^s_t . \bar{x}^s_t)$;

S^*_t "long-run" share of total cost incurred from factor input n in period t;

t time period;

T^s_m share of firm s total revenue earnt from output m $(p^s_m y^s_m / p^s . y^s)$;

u alternative subscript reference for time period;

u_{st} model residual in fixed effects cost function equation for firm s in period t;

v alternative subscript reference for time period;

w vector of factor input prices $(w_1, w_2, ..., w_N)$;

w_t vector of factor input prices in period t $(w_{1t}, w_{2t}, ..., w_{Nt})$;

w^s vector of factor input prices for firm s $(w^s_1, w^s_2, ..., w^s_N)$;

w^s_t vector of factor input prices for firm s in period t $(w^s_{1t}, w^s_{2t}, ..., w^s_{Nt})$;

$\bar{w}_n$	factor input price of factor n averaged across all firms in sample;
$\tilde{w}_{nt}$	"normalised" input price for factor n in period t $(w^s_{nt}/\bar{w}_n)$;
w^+_t	vector of user prices for firm's factor inputs consistent with the firm's usual accounting practices for measuring capital stocks $(w^+_{1t},w^+_{2t},...,w^+_{Nt})$;
x	vector of factor inputs used by a firm $(x_1,x_2,...,x_N)$;
x_t	vector of factor inputs used by a firm in period t $(x_{1t},x_{2t},...,x_{Nt})$;
x^s	vector of factor inputs used by firm s $(x^s_1,x^s_2,...,x^s_N)$;
x^s_t	vector of factor inputs used by firm s in period t $(x^s_{1t},x^s_{2t},...,x^s_{Nt})$;
$\bar{x}_n$	factor input n averaged across all firms in sample;
$\tilde{x}^s_{nt}$	"normalised" factor input n $(x^s_{nt}/\bar{x}_n)$;
x^+_t	vector of factor inputs measured by firm's usual accounting practices $(x^+_{1t},x^+_{2t},...,x^+_{Nt})$;
$X^s(y)$	input requirement set for firm s to produce output y;
y	vector of outputs produced by a firm $(y_1,y_2,...,y_M)$;
y_t	vector of outputs produced by a firm in period t $(y_{1t},y_{2t},...,y_{Mt})$;
y^s_t	vector of outputs produced by firm s $(y^s_1,y^s_2,...,y^s_M)$;
y^s_t	vector of outputs produced by firm s in period t $(y^s_{1t},y^s_{2t},...,y^s_{Mt})$;

$\bar{y}_m$ output m averaged across all firms in sample;

$\tilde{y}_{mt}$ "normalised" output m ($y^s_{mt}/\bar{y}_m$);

z_t quasi-fixed input for firm in period t;

δ_{1n} deprecation rate assumed in firm's accounts for factor input n;

δ_{2n} actual depreciation rate for factor input n;

Δ_1 ($n\times n$) diagonal matrix containing depreciation rates assumed in firm's accounts for each factor input ($\delta_{11},\delta_{12},\ldots,\delta_{1N}$);

Δ_2 ($n\times n$) diagonal matrix containing actual depreciation rates for each factor input ($\delta_{21},\delta_{22},\ldots,\delta_{2N}$);

ε^s degree of local returns to scale for firm s;

ε^s_t degree of local returns to scale for firm s in period t;

$\gamma(x_t,x_{t-1})$ adjustment cost function for change from factor input x_{t-1} to x_t;

γ_t adjustment cost factor in period t;

η_{gn} cross-price elasticity of the factor demand for input g with respect to the price of input n;

η_{gnt} cross-price elasticity of the factor demand for input g with respect to the price of input n in period t;

μ_t Lagrange Multiplier in cost minimisation problem for firm in period t;

π_n parameter in adjustment cost function for factor input n;

σ deflator used to define efficiency;

ζ_{1n} parameter in price expectations model for factor input *n*;

ζ_{2n} parameter in factor demand expectations model for factor input *n*;

REFERENCES

Aigner, D., Lovell, C. and Schmidt, P. (1977) "Formulation and Estimation of Stochastic Frontier Production Function Models", Journal of Econometrics, Vol. 6, p.21-37.

Anandalingam, G. and Kulatilaka, N. (1987) "Decomposing Production Efficiency into Technical, Allocative and Structural Components", Journal of the Royal Statistical Society, Series A, Vol. 150, p.143-151.

Apostolakis, B.E. (1988) "Translogarithmic Production and Cost Functions: A Synopsis", Economic Studies Quarterly, Vol. 39, p.41-63.

Banker, R., Charnes, A., Cooper, W. and Maindiratta, A. (1988) "A Comparison of DEA and Translog Estimates of Production Frontiers Using Simulated Observations from a Known Technology", p.33-55, in Fare, R. and Dogramaci, A. (editors) "Applications of Modern Production Theory: Efficiency and Productivity", Kluwer Academic Publishers.

Barnes, P. (1985) "UK Building Societies - A Study of the Gains from Merger", Journal of Business & Accounting.

Barrow, M. and Wagstaff, A. (1989) "Efficiency Measurement in the Public Sector: an Appraisal", Fiscal Studies, Vol. 10, p.72-97.

Bartelsman, E.J. and Dhrymes, P.J. (1991) "Productivity Dynamics: US Manufacturing Plants, 1972-1986", Columbia University, Department of Economics, Discussion Paper Series No. 584.

Berndt, E. and Khaled, M. (1979) "Parametric Productivity Measurement and Choice Among Flexible Functional Forms", Journal of Political Economy, Vol. 37, p.1220-1245.

Berndt, E., Fuss, M. and Waverman, L. (1977) "Dynamic Models of the

Industrial Demand for Energy", Electrical Power Research Institute, Research Report no. EA-580.

Berndt, E., Fuss, M. and Waverman, L. (1980) "Empirical Analysis of Dynamic Adjustment Models of the Demand for Energy in U.S. Manufacturing Industries, 1947-74", Final Report, Electric Power Research Institute.

Berndt, E. and Wood, D. (1979) "Engineering and Econometric Interpretations of Energy-Capital Complementarity", American Economic Review, Vol. 69, p.342-354.

Bond, S. and Devereux, M. (1990) "Economic Analysis and Company Accounts", Institute of Fiscal Studies, Working Paper No. W90/1.

British Telecommunications Engineering (1988) "Telecommunications in International Finance".

Callen, S. (1988) "Productivity, Scale Economies and Technical Change: Reconsidered", Southern Economic Journal, p.715-724.

Caves, D., Christensen, L., and Diewert, W. (1982a) "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity", Econometrica, Vol. 50, p.1393-1414.

Caves, D., Christensen, L., and Diewert, W. (1982b) "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers", Economic Journal, Vol. 92, p.73-86.

Caves, D., Christensen, L. and Tretheway, M. (1984) "Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ", Rand Journal of Economics, Vol. 15, p.471-489.

Charnes, A., and Cooper, W. (1985) "Preface to Topics in Data Envelopment Analysis", Annals of Operations Research, Vol. 2, p.59-94.

Charnes, A., Cooper, W. and Rhodes, E. (1978) "Measuring the

Efficiency of Decision-Making Units", European Journal of Operational Research, Vol. 2, p.429-444.

Clark, J.A. (1988) "Economies of Scale and Scope at Depoitory Financial Institutions: A Review of the Literature", Economic Review of the Federal Reserve Bank of Kansas City, p.16-33.

Cornwell, C., Schmidt, P. and Sickles, R. (1989) "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels", New York University, Economic Research Reports, Number 89/18.

Denison, E. (1979) "Accounting for Slower Economic Growth: the U.S. in the 1970's", Brookings Institute.

Denison, E. (1985) "Trends in American Economic Growth, 1929-1982", Brooking Institute.

Denny, M. and De Fontenay, A. (1988) "Cost Pressures and Relative Productivity in Canadian Telecommunications Firms", Chapter 5, p.131-148 in "Applications of Modern Production Theory: Efficiency and Productivity", Fare, R. and Dogramaci, A. (editors), Kluwer Academic Publishers.

Department of Trade and Industry (1991), License of British Telecommunications, Ammendments of 1991.

Diewert, W. (1976) "Exact and Superlative Index Numbers", Journal of Economterics, Vol. 4, p.114-145.

Diewert, W. (1981) "Duality Approaches to Microeconomic Theory", Chapter 12 in "The Handbook of Mathematical Economics", ed. Arrow, K. and Intriligator, M..

Drake, L. and Weyman-Jones, T.G. (1991) "Technical and Scale Efficiency in UK Building Societies", Loughborough University of Technology, Economic Research Paper No. 91/6.

Drake, L. and Weyman-Jones, T.G. (1992) "Productive and Allocative Inefficiencies in UK Building Societies: A Comparison of Non-Parametric and Stochastic Frontier Techniques", Loughborough University of Technology, Economic Research Paper, No. 92/2.

Epstein, L. (1981) "Duality Theory and Functional Forms for Dynamic Factor Demands", Review of Economic Studies, Vol. 48, p.81-95.

Epstein, L. and Yatchew, A. (1985) "The Empirical Determination of Technology and Expectations", Journal of Econometrics, Vol. 27, p.235-258.

Fare, R. and Dogramaci, A. (1988) "Applications of Modern Production Theory: Efficiency and Productivity", Kluwer Academic Publishers.

Fare, R. and Grosskopf, S. (1992) "Malmquist Productivity Indexes and Fisher Ideal Indexes", Economic Journal, Vol. 102, p.158-160.

Farrell, M. (1957) "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Series A, Vol. 120, p.253-266.

Ferrier, G. and Lovell, C. (1990) "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence", Journal of Econometrics, Special Issue, Vol. 46, p.229-245.

Fisher, I. (1952) "The Making of Index Numbers", Houghton Mifflin.

Forsund, F., Lovell, C. and Schmidt, P. (1980) "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement", Journal of Econometrics Supplement, Vol. 13, p.5-25.

Fuss, M. and McFadden, D. (1978) "Production Economics: A Dual Approach to Theory and Applications", North Holland.

Gollop, F. and Jorgenson, D. (1980) "U.S. Productivity Growth by Industry, 1947-73", Chapter 1, p.17-39, in Kendrick, J. and Vaccara, B. (editors) "New Developments in Productivity: Measurement and Analysis", Studies in Income and Wealth (Volume 44), University of Chicago Press, for the National Bureau of Economic Research.

Greene, W. (1980a) "Maximum Likelihood Estimation of Econometric Frontier Functions", Journal of Econometrics Supplement, Vol. 13, p.27-56.

Greene, W. (1980b) "On the Estimation of a Flexible Frontier Production Model", Journal of Econometrics Supplement, Vol. 13, p.101-115.

Gong, B. and Sickles, R. (1992) "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data", Journal of Econometrics, Vol. 51, p.259-284.

Grosskopf, S. (1986) "The Role of the Reference Technology in Measuring Productive Efficiency", Economic Journal, Vol.96, P.499-513.

International Data Corporation (1989).

Hadjimatheou, G. (1976) "Housing and Mortgage Markets: The UK Experience", Saxon House.

Hannan, T. and McDowell, J. (1987) "Rival Precedence and the Dynamics of Technology Adoption: an Empirical Analysis", Economica, Vol. 54, p.155-171.

Hardwick, P. (1989) "Economies of Scale in Building Societies", Applied Economics, Vol. 21, p.1291-1304.

Hunt, L.C. and Lynk, E.L. (1989) "Telecommunications Industry Divestiture in the UK: An Empirical Examination of the Case for Restructuring BT", University College of Swansea, Department of Economics Discussion Paper Series, No.89-07.

Jorgenson, D. and Griliches, Z. (1967) "The Explanation of Productivity Change", Review of Economic Studies, p.249-283.

Jorgenson, D. and Lau (1977) "Duality and Technology", Amsterdam: North Holland.

Kumbhakar, S.C. (1991a) "Estimation of Technical Inefficiency in Panel Data Models with Firm and Time Specific Effects", Economic Letters, Vol. 36, p.43-48.

Kumbhakar, S.C. (1991b) "The Measurement and Decomposition of Cost-inefficiency: The Translog Cost System", Oxford Economic Letters, Vol. 43, p.667-683.

Larson, B.A. (1992) "Technical Change, Input Substitution and the Specification of Dynamic Factor Demands", Economic Modelling, April 1992, p.129-136.

Lawrence, C. (1989) "Banking Costs, Generalised Functional Forms, and Estimation of Economies of Scale and Scope", Journal of Money, Credit and Banking, Vol. 21, p.368-379.

Lawrence, D. (1990) "An Adjustment-Costs Model of Export Supply and Import Demand", Journal of Econometrics, Vol. 46, p.381-398.

Lovell, C. and Schmidt, P. (1988) "A Comparison of Alternative Approaches to the Measurement of Productive Efficiency", Chapter 1, p.3-55 in Fare, R. and Dogramaci, A. (editors) "Applications of Modern Production Theory: Efficiency and Productivity", Kluwer Academic Publishers.

Madan, D.B. and Prucha, I.R. (1989) "A Note on the Estimation of Nonsymmetric Dynamic Factor Demand Models", Journal of Econometrics, Vol. 42, p.275-283.

Meeusen, W. and van den Broeck, J. (1977) "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", International Economic Review, Vol. 18, p.435-444.

Morrison, C.J. and Berndt, E.R. (1981) "Short-Run Labour Productivity in a Dynamic Model", Journal of Econometrics, Vol. 16, p.339-365.

Morrison, C.J. and Berndt, E.R. (1991) "Assessing the Productivity of Information Technology Equipment in US Manufacturing Industries", National Bureau of Economic Research, Working Paper No.3582.

Muellbauer, J. and Pashardes, P. (1988) "Tests of Dynamic Specification and Homogeneity in a Demand System", IFS Working Paper No. 88/6.

Nadiri, M.I. and Prucha, I.R. (1989) "Dynamic Factor Demand Models, Productivity Measurement, and Rates of Return: Theory and an empirical application to the US Bell System"

National and Provincial Building Society (1986), Annual Review for 1986.

Nickell, S. (1986) "Dynamic Model of Labour Demand", Chapter 9, p.473-522 in "Handbook of Labour Economics", Volume I, Ashenfelter, O. and Layard, R. (editors), Elsvier Science Publishers BV.

Nissim, J. (1982) "The Price Responsiveness of the Demand for Labour by Skill: British Mechanical Engineering: 1963-1978", Centre for Labour Economics, London School of Economics, Discussion Paper No.122.

Noulas, A., Ray, S. and Miller, S. (1990) "Returns to Scale and Input Substitution for Large U.S. Banks", Journal of Money, Credit and Banking, Vol. 22, p.94-98.

Office of Telecommunications (Oftel) (1992), "The Regulation of British Telecom's Prices", Consultative Document.

Panzar, J.C. (1989) "Technological Determinants of Firm and Industry Structure", Chapter 1, p.4-59, in "Handbook of Industrial Organisation", Volume I, Schmalensee, R. and Willig, R. (editors), North Holland.

Pashardes, P. (1986) "Myopic and Forward Looking Behaviour in a Dynamic Demand System", International Economic Review, Vol. 27, p.387-397.

Peterson, W. (1979) "Total Factor Productivity in the U.K.: A Disaggregated Analysis" in Patterson, W. and Schott (editors) "The Measurement of Capital".

Pindyck, R.S. and Rotemberg, J.J. (1983) "Dynamic Factor Demands and the Effects of Energy Price Shocks", American Economic Review, Vol. 73, No. 5, p.1066-1079.

Pruch, I.R. and Nadiri, M.I. (1986) "A Comparison of Alternative Methods for the Estimation of Dynamic Factor Demand Models Under Non-Static Expectations", Journal of Econometrics, Vol. 33, p.187-211.

Prucha, I.R. and Nadiri, M.I. (1991) "On the Specification of Accelerator Coefficients in Dynamic Factor Demand Models", Economic Letters, Vol. 1991, p.123-129.

Roach, S. (1987) "Technology and the Service Sector: America'a Hidden Competitive Challenge", Morgan Stanley.

Robertson, D. "Substitutability Between Labour Inputs in U.K. Manufacturing", Centre for Labour Economics, London School of Economics.

Roller, L. (1990) "Modelling Cost Structure: the Bell System Revisited", Applied Economics, Vol. 22, p.1661-1674.

Samuelson, P. and Swamy, S. (1974) "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis", American Economic Review.

Schmidt, P. (1984) "An Error Structure for System of Translog Cost and Share Equations", Econometrics Workshop Paper 8309, Michigan State University.

Sengupta, J. (1989) "Efficiency Analysis by Production Frontiers", Kluwer Academic Publishers.

Sickles, R. and Streitwieser, M. (1989) "The Structure of Technology Substitution and Productivity in the Interstate Natural Gas Transmission Industry under the Natural gas Policy Act of 1978", Economic Research Reports, New York University, Number 89/17.

Sickles, R. and Streitwieser, M. (1989) "Technical Inefficiency and Productive Decline in the US Interstate Natural Gas Pipeline Industry Under the Natural Gas Policy Act", New York University, Economic Policy Papers, Number 89/20.

Solow, R.M. (1957) "Technical Change and the Aggregate Production Function", Review of Economics and Statistics, p.312-320.

Spinnewyn, F. (1981) "Rational Habit Formation", European Economic Review, Vol. 15, p.91-109.

Stevenson, R. (1980) "Likelihood Functions for Generalized Stochastic Frontier Estimation", Journal of Econometrics Supplement, Vol. 13, p.57-66.

Timmer, C. (1971) "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency", Journal of Political Economy, Vol. 79, p.776-794.

Todd, D. (1985) "Productive Performance in West German Manufacturing Industry", Journal of Industrial Economics, Vol. 33, p.295-316.

Treadway, A. (1974) "The Globally Optimal Flexible Accelerator", Journal of Economic Theory, Vol. 7. p.17-39.

Varian, H. (1978) "Microeconomic Analysis", Norton.