Modeling and Forecasting International Credit Risk: The Case of Sovereign Loans

ELENA KALOTYCHOU

A thesis submitted for the degree of Doctor of Philosophy (Ph.D.) in Finance

CITY UNIVERSITY
CASS BUSINESS SCHOOL
FACULTY OF FINANCE

LONDON 2004

CONTENTS

	List of Tables	iv
	List of Figures	v
	Acknowledgements	vi
	Declaration	vii
	Abstract	viii
Chapter 1	**Introduction**	
	1.1 Overview of the Research	1
	1.2 Importance of Sovereign Risk: The Stakeholders	3
	1.2.1 The Concepts of Sovereign and Transfer Risks	3
	1.2.2 Policymakers and the Emerging Markets Case	4
	1.2.3 Financial Institutions and International Lending	6
	1.2.4 Regulators and the New Basel Capital Accord	7
	1.3 Motivation and Objective of the Thesis	8
	1.4 Layout of the Thesis	13
Chapter 2	Modeling Sovereign Debt Crises Using Panels	
	2.1 Introduction	15
	2.2 Literature Review	18
	2.2.1 Empirical Evidence on Country Default Risk	18
	2.2.2 Heterogeneity in Debt Repayment	20
	2.2.3 Evaluating Event Forecasts	23
	2.3 The Dataset	26
	2.4 Models and Estimation	29
	2.4.1 Country-specific Heterogeneity	29
	2.4.2 Region-specific Heterogeneity	31
	2.4.3 Time-specific Heterogeneity	32
	2.4.4 Time-varying Country Heterogeneity	33
	2.4.5 Inference-based Metrics for Model Selection	33
	2.5 Forecast Framework	34
	2.5.1 Statistical Loss Functions	36
	2.5.2 Economic Loss Functions	36
	2.5.3 Forecast Accuracy Tests	37

2.6 Results and Discussion
2.6.1 Inference-based Comparison
2.6.2 Out-of-sample Forecast Comparison
2.7 Conclusion

Chapter 3 **Inferences on the Determinants of Default and the Autocorrelation Problem**
3.1 Introduction
3.2 Data and Estimation
3.3 Serial Correlation Problem and Correction
3.4 Results and Discussion
3.5 Conclusion

Chapter 4 **Towards the Optimal Design of an Early Warning System for Debt Crises**
4.1 Introduction
4.2 Elements in the Design of an Optimal EWS
4.2.1 Warning Horizon for a Crisis Signal
4.2.2 Cut-off Probability Rate
4.2.3 Decision-maker's Loss Function
4.2.4 Forecast Combining Schemes
4.3 Methodology
4.3.1 Competing Classifiers
4.3.2 Combining the forecasts from LOGIT-M, LOGIT-R and K-clustering
4.4 The Dataset
4.5 Optimal Calibration of Forecasting Tools
4.5.1 Balancing the missed defaults and the false alarms
4.5.2 Well-behaved Loss Functions
4.5.3 Optimal Cut-off and Warning Horizon Combination
4.5.4 Optimal Number of Clusters
4.6 Optimal Forecast Combination
4.7 Conclusion

Chapter 5 **A Comparative Analysis of Sovereign Credit Migration Estimators**
5.1 Introduction
5.2 Literature Review
5.2.1 Caveats in Rating Migration Analysis
5.2.2 Accuracy of the Transition Probability Estimates
5.2.3 Non-Markov Effects and Time Variation in the Migration Process
5.3 Methodology
5.3.1 The Time-homogeneous Continuous Markov Chain Model

		5.3.2 Discrete Multinomial (DM) Estimator	118
		5.3.3 Continuous Hazard Rate Estimators	119
		5.3.4 Transition Horizon	125
		5.3.5 Measuring Overall Rating Mobility	126
		5.3.6 Bootstrapping the Rating Migration Process	128
	5.4	The Dataset	130
	5.5	Properties of Migration Matrix Estimates	134
		5.5.1 Discrete versus Continuous Estimators	134
		5.5.2 Increasing the Transition Horizon	139
		5.5.3 The Time-homogeneity Assumption	140
		5.5.4 Robustness Checks	143
	5.6	Detecting Non-Markovian Behaviour	144
		5.6.1 Testing for Markov Structure of the Transition Matrix	145
		5.6.2 Testing the Duration Effects Hypothesis	146
		5.6.3 Testing the Rating Momentum Hypothesis	148
	5.7	Conclusion	149
Chapter 6	**Conclusion**		
	6.1	Concluding Remarks	167
	6.2	Further Research	170
	References		173
	Appendices		
	Appendix 2.1:	Country Composition and Indicators	191
	Appendix 2.2:	Historical Sovereign Defaults per Country, 1984-2002	192
	Appendix 2.3:	The Cross-Validation (Jacknife) Method	195
	Appendix 2.4:	Definition of Selected Variables	196
	Appendix 2.5:	Expected Impact of Selected Variables	197
	Appendix 2.6:	Cross-Section and Regional Estimates	202
	Appendix 4.1:	Clustering by K-Means Algorithm	203
	Appendix 4.2:	Historical Sovereign Defaults per Year, 1984-2002	204
	Appendix 4.3:	Logit Estimates 1984-1995	205
	Appendix 5.1:	Bootstrapping the Ratings Migration Process	206
	Appendix 5.2:	Moody's Foreign-Currency Sovereign Bond Issuers	207
	Appendix 5.3:	Moody's Ratings Distribution	209
	Appendix 5.4:	Summary Statistics for Sovereign Rating Durations	211

List of Tables

Table 2.1:	Regressor Set Selected for World and Regional PLOGIT Models	51
Table 2.2:	Model Comparison on the Basis of Information Criteria	52
Table 2.3:	Statistical Significance of Country-specific Effects	53
Table 2.4:	Statistical Significance of Time- and Regional-specific Effects	54
Table 2.5:	Parameter Estimates of Default Probability Models, 1984-2000	55
Table 2.6:	Out-of-sample Forecast Analysis	56
Table 3.1:	Logit EWS Models for Sovereign Default, 1984-2000 (QML Standard Errors)	67
Table 3.2:	Logit EWS Models for Sovereign Default, 1984-2000 (h.a.c Standard Errors)	78
Table 4.1:	Optimal Calibration of Classifiers over Estimation Window	94
Table 4.2:	Out-of-sample Forecast Ability of Competing Classifiers	95
Table 4.3:	Stability of Forecast Ranking over Holdout Period	96
Table 4.4:	Combination Weights Based on KK-logit Regression	97
Table 4.5:	Forecast Ability of Combined Classifiers	98
Table 4.6:	Ratio of the Classifier's Loss to the Loss of the Naive Predictor	99
Table 5.1:	One-year Homogeneous Rating Transition Probability Estimates	152
Table 5.2:	Summary Statistics for Bootstrap 1-year Default Probability Estimates	153
Table 5.3:	Bootstrap Tests for Differences between Discrete and Continuous Estimator	154
Table 5.4:	Longer Horizon Homogeneous Transition Probability Matrices	155
Table 5.5:	Longer Horizon Heterogeneous Transition Probability Matrices	157
Table 5.6:	Summary Statistics for Bootstrap 2-year Default Probability Estimates	158
Table 5.7:	Bootstrap Tests for Differences between Homogeneous and Heterogeneous Estimator	159
Table 5.8:	Logit Test Results for Non-Markov Effects	160

List of Figures

Figure 4.1:	The Type I and Type II errors and the Cut-off Rate	100
Figure 4.2:	The Overall Loss for Different Cut-off Rate and Warning Horizon Combinations (LOGIT-M)	101
Figure 4.3:	The Cut-off Rate and the Warning Horizon	102
Figure 4.4:	Calibration of the Number of Clusters	103
Figure 4.5:	Overall Loss for Individual and Combined Forecasts	104
Figure 5.1:	Aggregate Ratings Distribution, 1981-2003 (Moody's)	161
Figure 5.2:	Discrete versus Continuous Bootstrap Default Probability Estimates (high credit-quality ratings)	162
Figure 5.3:	Discrete versus Continuous Bootstrap Default Probability Estimates (low credit-quality ratings)	163
Figure 5.4:	Homogeneous versus Heterogeneous Bootstrap Default Probability Estimates (high credit-quality ratings)	164
Figure 5.5:	Homogeneous versus Heterogeneous Bootstrap Default Probability Estimates (low credit-quality ratings)	165
Figure 5.6:	Eigenvalues and Transition Horizon	166
Figure 5.7:	Second Eigenvector for Different Transition Horizons	166

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to my supervisor, Dr Ana-Maria Fuertes, for her invaluable guidance and continuous encouragement throughout the course of the Ph.D. I am grateful to Cass Business School for sponsoring me throughout the duration of my studies and to the people that helped to make this possible.

This thesis has benefited from valuable feedback at various presentations. I would like to thank participants at the 2004 Meeting of the *Society for Computational Economics* on Computing in Economics and Finance at the University of Amsterdam, the 2004 *European Economic Association* and *Econometric Society European Meeting* at Universidad Carlos III, Madrid, the 2004 *Multinational Finance Society* Conference in Istanbul and the 2004 *Australasian Finance and Banking* Conference at the University of New South Wales, Sydney, as well as seminar participants at Cambridge University and Cass Business School. In particular, I appreciate the useful suggestions and comments from Frank Packer of the Bank of International Settlements. I am also grateful to Alessio Balduini of Moody's Investors Services for providing the dataset.

Special thanks to my colleagues and fine friends Dr Sotiris K. Staikouras, Luigi Colombo and Marwa Khalaf-Allah for helpful discussions, comments and so much more during this period of my life. I would also like to thank the Head of the Finance Faculty at Cass Business School Professor Alec Chrystal, the Ph.D. Director Professor Giovanni Urga and the Ph.D. secretary Margaret Busgith for her patience and help.

Last, but not least, special thanks are due to my parents and my brothers Marios and Panayiotis, for standing by me throughout. Without their love and support this thesis would not have come into existence.

Elena Kalotychou
London 2004

Declaration

The author of this thesis grants power of discretion to the University Library to allow this thesis to be copied in whole or in part without further reference to her. This permission covers only single copies made for study or research purposes, subject to normal conditions of acknowledgement.

Abstract

This thesis investigates the relative merits of econometric modeling, statistical and judgmental techniques for predicting debt crises and assessing the risk of credit migration. The increased reliance on econometric or statistical approaches and credit rating systems in risk management has intensified the need for more rigorous analysis of their finite sample properties. A better understanding of the available tools has implications for credit risk management, regulation and policy decision-making.

The thesis contributes to the extant sovereign risk literature in three areas. First, it addresses the question of whether controlling for unobserved heterogeneity is important for predicting debt crises and explores a pervasive inference problem in Early Warning Systems (EWSs). Second, it addresses the development of an 'optimal' EWS for sovereign debt crises that accommodates the decision maker's preferences. Third, it considers the measurement of sovereign credit migration matrices using different estimators and explores non Markov effects in the rating dynamics.

Chapter 2 confronts competing models of sovereign default that differ in how country-, region- and time-specific effects are treated. Statistical tests and information criteria overwhelmingly favour more complex models with country heterogeneity that possibly changes over time. However, simplicity beats complexity in terms of forecasting. Simple pooled logit parameterizations that control either for regional heterogeneity or for time effects produce the most accurate forecasts and outperform several naive predictors.

Chapter 3 investigates the severity of the autocorrelation problem in EWS of sovereign default. This stems from seeking to provide crisis warnings over a horizon that is longer than the frequency at which the forecasts are updated and from the sluggishness of the typical exogenous indicators. Neglecting residual serial autocorrelation in such models is shown to be far from innocuous. Inferences are overturned when using a correction. This phenomenon is generally clearer for the macroeconomic ratios that are more persistent.

Chapter 4 combines three fundamentally different classification techniques — econometric, statistical and judgmental— to produce an EWS for sovereign default. The optimal choice of crucial EWS elements is shown to depend on the decision-makers' preferences. The forecast ranking of classifiers is found to be unstable and overall the classifiers appear to have different strengths. Payoffs from forecast combination are documented and the combining scheme is shown to depend on the decision-makers' loss function.

Chapter 5 turns to the estimation of sovereign transition probability matrices and evaluates the popular discrete multinomial estimator against two continuous hazard rate methods that differ in their treatment of time-heterogeneity. Bootstrap simulations of the rating generating process reveal interesting insights. Hazard rate estimators yield more reliable default probabilities. Efficiency is further enhanced upon relaxing homogeneity. Downgrade momentum and duration effects are found to be present in the rating process.

CHAPTER 1: INTRODUCTION

1.1 Overview of the Research

Over the last two decades, several costly lessons have led to a new emphasis on international credit risk management. The key impulse for the renewed concern was the high-profile debt defaults by sovereigns. In many regions, through a spillover mechanism, various countries have experienced at least one serious bout of debt-servicing trouble in the recent past. In the early 1990s the Mexican peso crisis led to some debt defaults, while in 1997 the Asian 'meltdown' evidently spread through Eastern Europe and Latin America triggering a chain of unprecedented credit jitters — in 1998, the plummeting of the Russian ruble pushes the Russian government to default on the US dollar denominated 'MinFin' bonds; 1999, Pakistan seeks a restructuring of its foreign bank debt obligations with the Paris Club lenders;[1] 2000, Ukraine defaults on Eurobonds and US dollar denominated bonds; 2000, Ecuador declares the first-ever default on Brady Bonds;[2] 2001, Argentina announces the largest sovereign default in history. In the aftermath of the upheaval, financial institutions and regulators became painfully aware of the credit exposure to international lending. Multinational banks that had lent heavily in these regions over the last two decades failed in record numbers. As the Asian markets collapsed

[1] The Paris Club is the term used to describe the monthly meetings between 19 creditor and debtor nations to negotiate and discuss each country's financial obligations to the others. Frequently, relief measures are drawn up for third world and developing countries that cannot meet their liabilities. The Paris Club permanent creditor members are governments with large claims on various other governments throughout the world (the claims may be held directly by the government or through its appropriate institutions). They include European countries, as well as USA, Japan, Australia and Canada.

[2] Brady bonds are dollar bonds issued by a country to be swapped for outstanding loans. They have a much longer maturity than that promised on the original loans and a lower coupon than the interest rate on the original loan. The principal is usually collateralized through the holding of U.S. T-bonds. Once loans are swapped for bonds by banks and other financial institutions, they can be sold on the secondary market. This restructuring program was developed under the authority of the U.S. Treasury's 1989 Brady Plan and other international organizations such as the International Monetary Fund (IMF).

financial institutions in Japan and Hong-Kong failed and were either forced to merge or reconstruct. In 1998, Chase Manhattan, which has been the biggest lender in Latin America, took off its books $2.6 billion aiming to reduce its credit exposure to the continent. As a result, country risk measurement is at the centre of modern risk management and international capital allocation. Its importance has been recently emphasised by Kristin Forbes (MIT, US Treasury): "What we should care about and what I would like to see more work go into is models predicting things such as external financing difficulties and financial systems vulnerabilities, just as a few examples".[3]

The objective of this thesis is to explore issues pertaining to modeling, estimation and forecasting in the context of sovereign debt and to investigate how such issues can be exploited in designing a sovereign risk assessment device. A number of features make the current study distinct. First, a systematic analysis of the importance of cross-country, regional and timely differences in sovereign repayment performance is carried out. Second, a rigorous forecast evaluation framework for sovereign default is developed, which could also be applied in the related contexts of currency and banking crises. Third, we delve into a number of empirical issues regarding the optimal design of an Early Warning System (EWS). Fourth, we compare the finite sample properties of rival transition matrix estimators on which very little is known in the context of sovereigns. The dataset for the analysis includes sovereign debt and credit rating data for a broad range of developed, emerging and less developed countries (LDC) since the early 1980s. The data span encompasses the major recent sovereign debt crises in Latin America, Asia and Eastern Europe. The findings from this study will be useful for practitioners in Central Banks, financial institutions and rating agencies interested in the development of EWSs for sovereign default and the estimation of sovereign rating migration probabilities.

The remainder of this chapter is structured as follows. Section 1.2 discusses the concept of sovereign risk and its importance. Section 1.3 describes the motivation and objective

[3]Panelist in the '2001 Economic Forum' organized by the IMF.

of this study. Finally, Section 1.4 outlines the layout of the thesis.

1.2 Importance of Sovereign Risk: The Stakeholders

1.2.1 The Concepts of Sovereign and Transfer Risks

A commonly used definition of sovereign risk provided by Claessens and Embrechts (2002) is: "The risk of exposure to losses caused by events in a particular country, which may be under the control of the government but not under the control of a private enterprise or individual". The notion of sovereign risk is very broad and encompasses all forms of foreign lending in a country — to the government, a bank, a private enterprise or an individual. It comprises a variety of economic, political, social, cultural or legal factors that could render the fulfilment of foreign currency debt obligations uncertain.

Sovereign risk encapsulates private credit risk. A distressed government may impose prohibitive exchange restrictions to private foreign borrowers within its domicile so that the latter cannot make any debt repayment despite being in a good credit position. This is an example of transfer risk and could be one of the most important drivers of country risk. Thus, international lenders to private entities are also concerned with the sovereign risk quality of the country in which the private borrower resides. The latter is reflected in the risk weights of the New Basel Accord whose implications are described in detail below in Section 1.2.4.[4] Another feature that differentiates sovereign risk analysis from corporate credit risk analysis is the important role of a sovereign's *willingness* to fulfill debt obligations. By contrast, corporate credit risk revolves around the *ability* for debt repayment, which is an easier to quantify concept. As opposed to corporates, sovereigns cannot file for bankruptcy. As a result, the characterisation of a sovereign default event has been the subject of various controversies. There is no single definition of sovereign

[4]Risk weighting for interbank loans includes two options. The first is based on the (rating-based) weighting of the sovereign in which the bank is incorporated. The second is based on the assessment of the individual bank.

default, however, the leading rating agencies have set the grounds for a definition that is becoming commonly accepted and utilised. According to Moody's (Moody's, 2003), a sovereign issuer is in default if one or more of the following conditions are met:

1. There is a missed or delayed payment of interest and/or principal, even if the delayed payment is made within the 'grace period', if any.[5]

2. A distressed exchange occurs (rescheduling or restructuring) where:

   - The issuer offers lenders a new security or package of securities that amount to a diminished financial obligation such as debt instruments with a lower coupon or par value and/or longer maturity.

   - The exchange had the purpose of aiding the borrower to avoid a stronger event of default (such as missed interest or principal payment).

Such rescheduling agreements are reached after negotiations between the sovereign borrower and international loan syndicates that comprise major international banks.[6]

1.2.2 Policymakers and the Emerging Markets Case

The growing degree of interdependence between emerging and developed markets has also had implications for country risk assessment. Emerging markets have increased their borrowing considerably in order to modernize their economies and improve their competitiveness in western markets. The degrees of liberalization of convertible-currency imports and borrowing in the individual countries differed depending upon internal political choices. Latin American countries, in particular, have had a long history of reliance on foreign capital to finance their developments. In fact, the debt crisis of 1982 was born

[5] 'Grace period' is the period between the rescheduling and when the sovereign starts servicing the rescheduled debt obligation.

[6] Citigroup, for example, was chosen as the lead bank negotiator in five major loan reschedulings in the 1980s, as well as in both the Mexican and South Korean reschedulings in the 1990s.

there, and a significant amount of outstanding Brady bonds is still associated with Latin American borrowers.

Sovereign default risk plays a prominent role for those countries' economic, social and political well-being. Inability to generate the required foreign exchange will lead to balance of payments disequilibrium. Restoration of equilibrium implies economic adjustments like relative price changes, resource reallocation and income redistribution that are costly for a country's growth, unemployment and standards of living. Further, three other costs are inflicted by sovereign debt defaults. The first is associated to the cost of losing access to international capital markets (Eaton and Gersowitz, 1981). Investors have embraced emerging markets in the mid 1990s, however, the recent turmoil in some of them was enough to lead the total foreign direct investment flows downsurging by the end of the decade. The second refers to the costs due to direct sanctions such as the elimination of trade credits or the seizure of assets (Bulow and Rogoff, 1989). The third concerns the costs associated with a domestic stock market downhill. Recent anecdotal evidence suggests that country default risk has profound implications for a country's stock market performance (Clark and Kassimatis, 2004). Between December 1994 and February 1995, during the Mexican peso crisis, the Mexican stock market index fell by 38.7%. Asian stock markets plunged during the Asian crisis.[7] In August 1998, the Russia watched its stock market plummeting by 41.3% in the first month of the Russian crisis. The withdrawal of foreign capital from the domestic market can exacerbate the balance of payments disequilibrium and make economic adjustment even more costly.

Governments of emerging markets today have greater incentives to avoid such penalties as their future depends heavily on their credit evaluation record. For emerging markets that rely on foreign lending to finance their needs, these are important issues as a troubled external financial profile could jeopardise their development.

[7] From July 1997 to February 1998, the stock market in Thailand fell by 48.4%, in Indonesia by 81.7%, in Malaysia by 58.4%, in Philippines by 49.2% and in Korea by 63.1%.

1.2.3 Financial Institutions and International Lending

Among the various risks that multinational banks face in modern financial markets is the financial stability of a borrower. Banks are seriously affected, through their loans, by worldwide economic changes and energy prices. In recent years, such exposure has caused enormous problems for US banks lending to LDC, Latin America and Asia (Saunders and Cornett, 2003). The timely fulfillment of debt obligations is not a modern phenomenon (Avramovic, 1958). However, over the last two decades, the magnitude of repayment problems has reached unprecedented levels and has increasingly involved commercial banks. In 1982, when Mexico and Brazil announced their debt moratoria, 80% of the US banks' sovereign exposure was concentrated in Latin America. Various banks were later forced to increase their loan loss reserves with Citicorp alone setting aside $3 billion. One could easily argue that in many cases these loans appear to have been made with little judgment regarding the creditworthiness of the sovereign country.

One lesson from World War II was that economic distress often leads to political turmoil, international tensions and military conflict. Factors that trigger sovereign risk could be attributed to various and complicated dynamics such as an economic decline, social unrest, possibility of war or a change in political ideology. Such uncertainties are clearly illustrated by the rapid global changes that followed the collapse of Eastern Europe during the last decade. Despite great efforts in assessing and measuring a country's risk, financial institutions have been caught off guard by sudden unexpected changes in a country's economic climate. The recent financial crises in Mexico (1994-95), Asian (1997-99) and Latin American (1999-01) have raised awareness on the importance of country risk. Only at this time US financial institutions, armed with the experience of the 1980s, limited their exposure to one third of that of their counterparts in Europe and Japan. Improved sovereign risk assessment techniques and mechanisms in dealing with such exposure did play an important role (Saunders and Cornett, 2003). Nevertheless, J.P.Morgan Chase, Bank of America and other major players announced losses from emerging markets amounting to millions of dollars in the late 1990s.

The importance of sovereign risk analysis is thus perceptible in various fronts. First, it is necessary to monitor the performance of existing loans and other investments (e.g. debt and equity claims). Second, organizations such as the IMF and the World Bank would be able to prevent crisis and effectively support countries exhibiting signs of financial instability. Third, the effects of a financial crisis are not only felt in emerging markets but strong economies might suffer 'equally' well by such turmoil. Fourth, sovereign risk analysis is not a tool for solely predicting financial crises; it is a vehicle of improving the decision-making process regarding capital budgeting and/or financing issues. Fifth, serious banking problems (or even failures) in developed economies, due to international lending, are presumed to generate serious negative externalities. Finally, due to integration of financial markets it is sensible that foreign direct investment in LDC is necessary if global markets are to prosper.

1.2.4 Regulators and the New Basel Capital Accord

Financial institutions have devoted many resources to developing internal models for better quantifying their risks and assigning sovereign risk capital. These efforts have been recognized and encouraged by bank regulators. In response to that, in January 2001 the proposal for a New Basel Capital Accord (Basel II) launched by the Basel Committee on Banking Supervision, has reinforced the interest in obtaining loan default probabilities and rating transition matrices.[8] Banks are expected to apply the new risk management concepts not only to domestic borrowers but also to international lending in the next few years. In the light of Basel II, sovereign ratings by rating agencies, as well as internal bank ratings are expected to play an important role in the measurement of credit risk.

[8]See Basel Committee on Banking Supervision (2001). The Basel Committee is a regulatory body under the wings of the Bank of International Settlements.

One of the most important goals behind Basel II is to modify the actual risk weights using a methodology that more clearly identifies the differences in risks of various instruments. Both practitioners and academics already predict that rating systems and credit migration matrices will play a larger role in capturing these differences. Credit migration matrices, which characterise the expected changes in credit quality of obligors are already cardinal inputs to many risk management applications, including portfolio risk management, modeling the term structure of credit spreads and pricing credit derivatives. Within the Basel II framework, two approaches are allowed for measuring credit risk: the standardized approach and the Internal Ratings Based (IRB) approach. The standardised approach is a refinement of Basel I and postulates the use of ratings from the rating agencies for assessing the economic risk capital.[9] The IRB method gives flexibility to banks to use their internal risk management systems for calculating capital requirements. Specifically, the Basel Committee advocates statistical models that capture all key variables driving credit risk, as well as expert (judgmental) risk ranking as tools for determining the overall risk level of credit instruments or portfolios.[10] Basel II leaves ample room for a statistical approach to risk.

1.3 Motivation and Objective of the Thesis

This study empirically investigates the properties of some of the available tools for assessing sovereign default risk and provides insights on their relative ability to predict debt crisis episodes and quantify the likelihood of sovereign credit rating migration. For this

[9]In Basel II, the sovereign creditworthiness risk weights for speculative grade rated debt instruments are between 100% and 150%, implying that banks have to set aside capital equal to 8% and 12% of such loans respectively. Within Basel II, risk weights are based on credit ratings. By contrast, the 1988 Capital Accord (Basel I) provided only three weights for capital requirements based on the OECD/non-OECD distinction. Those weights were 0% for OECD sovereign bonds, 20% for all claims on OECD banks and short-term claims on non-OECD banks, and 100% for long-term claims on banks, corporates and sovereigns of non-OECD countries. These weights correspond to capital requirements of 0%, 1.6% and 8%, respectively. (cf: Monfort and Mulder, 2000, p. 3 and 5).

[10]Basel Committee on Banking Supervision (2001; p.51, paragraph no. 266).

purpose, it utilizes panel discrete-choice models, non-parametric classification tools, sovereign credit rating systems as well as migration probability estimators based on survival theory. The importance of international credit risk for risk managers and market participants together with the existing gaps in the empirical literature constitute the primary motivation of the thesis.

The empirical literature on sovereign debt and the determinants of default mostly rests on panel probit or logit models. Since sovereign defaults on bonds are only recent, the extant studies mostly include default events on bank loans, which upsurged between the onset of the 1980s Latin American debt crisis and the early 1990s. In this context, researchers have to deal with the task of explaining and predicting a relatively rare, but persistent event. Moreover, the available emerging and developing country indicators for such studies are typically annual series over 10 to 20 years, often with missing observations, which make individual country analyses unfeasible. The literature has side-stepped this problem by exploiting the cross-sectional variation in country default experiences to increase the power of the models. Panel logit and to a lesser extent simple cross-section formulations are thus typical in the literature. The majority of panel studies impose homogeneity both across countries and over time.

Restrictive full homogeneity may incur two sources of bias. First, estimates can be substantially biased if country-specific behavior is not accounted for. The diversity of the countries employed in order to have a reasonable sample size, casts doubt on the country-homogeneity assumption. By assuming that the responses to common systematic factors in different countries are similar, one ignores any unsystematic country-specific dimensions of risk attributable to political, social, religious aspects, as well as to the diverse economic policies pursued in different countries. Attempts to quantify such subjective factors were made by Balkan (1992), Brewer and Rivoli (1997) and Haque et al. (1997) by creating variables of political risk, such as level of democracy and political instability. However, research in this area is scarce. Second, homogeneity over time ignores the rapidly changing nature of emerging and developing economies. Both issues have raised concern within the academic circles. Despite its obvious importance, there are no studies that deal directly with the impact of neglecting heterogeneity, for instance, on predictive performance. Motivated by the work of Baltagi and Griffin (1997) and Baltagi et al. (2000) in the context of linear panel models we explicitly address the question of whether accounting for latent heterogeneity improves sovereign default models. For this purpose, we rely both on statistical and forecasting criteria. The analysis seeks to reconcile the empirical and anecdotal evidence regarding the key role of country, regional or time-specific effects, as well as unify and extend the previous literature.

A number of other important issues in the empirical literature on sovereign default warrant further investigation. First, scant attention has been paid to the 'optimal' design of an EWS for sovereign default tailored to the decision-maker's preferences. It has been argued that the decision maker's loss specification can fundamentally determine a number of input parameters associated with the development of EWSs (Bussière and Fratzscher, 2002). Second, the assessment, comparison and combination of competing forecasting models has received superficial attention, if any. Third, besides probit and logit models, other statistical classification techniques have been exploited in the literature. In addition, external sovereign credit ratings provided by leading rating agencies (Standard and Poor's and Moody's) and internal ratings from major international banks (Institutional Investor) have also been considered.

So far no consensus has been reached about the 'best' classification method and information set for the development of an EWS for sovereign default. There is only very limited evidence in this regard and the extant studies are subject to the above criticisms regarding forecast evaluation. Our work aims to provide a rigorous assessment of forecasting ability in order to provide insights regarding the relative strengths and weaknesses across: a) homogeneous and heterogeneous panel logit models and b) other fundamentally different approaches based on non-parametric statistical classification and credit ratings. Our framework embodies both statistical and economic criteria, predictive performance tests and reasonable naive benchmarks. In the present context, the latter may be hard to beat due to the inherent persistence of the external debt repayment behaviour. Further, our framework emphasizes the key role of the decision maker's loss function in model calibration and in choosing among rival combination schemes.

Most of the studies that adopt probit or logit models to develop EWSs for sovereign debt crises are vulnerable to shortcomings that will invalidate inferences. The EWS literature generally seeks to forecast crises over a multiple period horizon (i.e. several months or years ahead), which is typically longer than the frequency with which the forecasts are updated. The idea is to provide earlier warning signals while, at the same time, make use of the most recent data available in updating forecasts. However, this comes at the expense of overlapping forecasts that induce moving average prediction errors. The problem is compounded if serial correlation is also present in the explanatory variables, a stylised fact of macroeconomic and financial ratios. For nonlinear models such as logit or probit, this does not invalidate the consistency of the estimators, provided no other misspecification is present (Poirier and Ruud, 1988; Gourieroux et al., 1984). The standard errors, however, will be biased and tests based on them will have incorrect size. Little attention has been paid by the extant literature to this issue. We document the magnitude of the autocorrelation problem in the context of sovereign default and implement a correction.

While it is important to investigate the characterisation and prediction of sovereign default events for risk assessment and policymaking purposes, the broader sovereign rating migration is at least as important and merits particular attention. Reports on rating migrations published by rating agencies are used all around the globe by diverse practitioners, such as sovereign bond investors or analysts and traders working in the fastly growing credit derivatives market. Credit migration matrices are intimately linked to modern risk management tools such as Value-At-Risk (VaR) analysis and credit portfolio models and to the determination of regulatory capital complying with the Basel II rules. Their adequate estimation is thus quite important.

There are some stylised facts in the context of corporate credit migration, but relatively little is known in the context of sovereigns. First, corporate credit migration does not conform to the Markov behaviour (Bangia et al., 2002; Lando and Skodeberg, 2002; Kavvathas, 2001). Under the Markov assumption, the future realization of the rating only depends on its previous value. Violation of this implies that the history of the rating process carries information about its future, and thus about future transition probabilities. The non-Markov properties reported in the literature can be classified in three types. One is rating momentum or the influence of the past rating migrations on the current migration probability.[11] The second refers to the dependence of transition probabilities on rating durations, that is, on the time spent in each rating. The third concerns the dependence between rating migrations and the state of the business cycle. Assuming a homogeneous Markov process to develop migration probability measures implies ruling out explicit time and state dependence in the migration process, as well as business cycle effects.

The literature has not investigated these issues in the context of sovereign migrations as yet. The bare estimation of sovereign migration matrices has received scant attention. The corporate and sovereign migration contexts are different in several respects. First, the available samples of sovereign ratings are relatively small, both in terms of time series length (rating agencies started rating emerging markets in the early 1990s) and cross-sectional dimension. Second, under the definitions of default employed by rating agencies there are few default cases on foreign-currency rated bonds. Against this background, it is important to analyze the reliability of the transition matrices that have been provided

[11] Sometimes the term 'rating drift' is used alternatively.

by rating agencies in the recent years. By exploiting the complete Moody's database from 1981 to 2004, we contribute to the literature by comparing different sovereign migration measures. We examine the extent to which continuous hazard rate estimators have better properties than the discrete multinomial estimator (or cohort method). The latter can be considered as the industry standard and so it would be interesting to see how it compares with more sophisticated survival analysis methods. A second objective is to assess how important it is to relax the time homogeneity assumption in the context of sovereign transitions. Finally, we test for two non-Markovian effects that in rating transitions.

1.4 Layout of the Thesis

The remainder of the thesis consists of four empirical essays regarding the analysis of sovereign risk. The thesis can be divided into two parts. Chapters 2, 3 and 4 examine the characterisation and prediction of sovereign default events within the context of an optimal EWS. The second part deals with the estimation of sovereign transition matrices and related issues and is presented in Chapter 5.

Chapter 2 compares rival models of sovereign default that differ in how country-, region- and time-specific effects are treated. The analysis is based on a diverse panel of countries from Africa, Asia, East Europe, Latin America and the Middle East over twenty years. The quality of the models is gauged by means of likelihood ratio and Hausman type tests, information criteria and judgements on the plausibility of the estimates. In addition, an out-of-sample forecast evaluation based on statistical- and economic-loss functions is conducted. Each model is benchmarked against simple forecasts. These include the random prediction implicit in the Pesaran-Timmermann test and the most-frequent-event prediction. A Diebold-Mariano test is deployed to compare the forecasts.

Chapter 3 addresses the residual autocirrelation problem of EWSs for financial crises that attempt to explain a forward-looking indicator which is persistent by construction.

The typical framework is a panel logit/probit regression estimated by maximum likelihood. The purpose of this chapter is to illustrate that one should not overlook the presence of residual serial autocorrelation in such models. To this end, different logit specifications for sovereign default are deployed. These range from a baseline pooled model to a random coefficients model. We document the extent of the problem and implement a correction following the Newey-West approach.

Chapter 4 utilizes three different classification methods — K-means clustering a logit regression based on macrovariables and a logit based on bankers' credit ratings — to develop an EWS for sovereign default. Specifically, the chapter illustrates how the optimal choice of parameters depends on the decision-makers' preferences. Next, it investigates the potential benefits of combining econometric, judgmental and non-parametric sovereign default forecasts derived from the three classification techniques. Combining forecasts of discrete-variables requires different approaches based either on logit regression or on voting-type prediction rules. In this context, the benefit from combination is not as clearcut, since the expected loss is not directly related to the forecast error variance.

Chapter 5 addresses the estimation of sovereign credit migration matrices and tests for non-Markov effects in the rating migration process. Three estimators are considered: discrete-time multinomial and two variants of continuous-time hazard rate — one imposing, the other relaxing time homogeneity. The transition matrices are compared on the basis of bootstrap simulations reflecting the continuous-time underlying rating migration process. The default probabilities associated with the three approaches are examined, along with the rating mobility inherent in each transition matrix as depicted by matrix norms. Finally, the presence of momentum and duration effects in the rating process is investigated through spectral and panel logit analyses.

Chapter 6 concludes the thesis by providing an overview of our research and a summary of the results. Finally, the chapter suggests potential directions for future research.

CHAPTER 2: Modeling Sovereign Debt Crises Using Panels

2.1 Introduction

The formulation and estimation of binary-choice models for panel data has been the subject of a rapidly growing literature.[1] Panels can provide insights which are not available in pure cross-section or time-series data (see Baltagi, 2002). The choice of estimator depends on whether the data are conceptualized as repeated cross-sections or a pool of time series. Pure cross-section estimators cannot allow for country heterogeneity. Pooled time series estimators can pick up differences in behavior across individuals not captured by the included regressors. Default on sovereign debt is not a frequent event for a given country and definitely not a short-term situation. As a result, the binary dependent variable representing the latter exhibits small variation. In addition, the available emerging/developing country indicators for such studies are typically annual (moderate T) series over 10 to 20 years, often with missing observations. The above two issues combine to make individual country analyses unfeasible. Researchers have circumvented this problem by jointly exploiting time series on typically 20 to 130 countries (large N panels). Over 80% of these studies employ logit models. Simple pooled logit regressions are common, although the validity of the implicit full homogeneity assumption has been questioned (McFadden et al. 1985, Hajivassiliou, 1987). As Schleifer (2003) puts it: "Sovereign debt markets could not be more different".

The limited number of studies that control for latent country heterogeneity use either fixed or random effects logits. One exception is Oral et al. (1992) who allow for fixed country effects both in the intercept and slope coefficients of the domestic signals. Some evidence in the related scenario of currency crises suggests that the relevant heterogeneity

[1] A comprehensive survey can be found in Arellano and Honoré (2001).

occurs at a broader, regional level (Burkart and Coudert, 2002; Staikouras, 2005). In the context of sovereign default, regional differences have been accommodated via regional fixed effects (Feder et al., 1981). Some studies find significant time effects using year dummies (Aylward and Thorne, 1998) or including global macroeconomic variables such as OECD economic growth (Lee, 1991; Detragiache and Spilimbergo, 2001). This stresses the importance of exogenous world shocks on default risk.

The large number of empirical models available have led to at best mixed evidence regarding the determinants of sovereign default. The estimates are based on different samples (countries and time span) and there is no unanimous default definition. Both of these aspects make the model comparison onerous. There is also the non-trivial issue of how to carry out the comparison. One can employ statistical hypothesis tests of parameter restrictions. Alternatively, the 'best' model can be chosen using extant model selection criteria. With a small number of degrees of freedom, this approach can lead to quite close values for the criteria and hence, to model selection instability. A third common approach is to compare the plausibility of the estimates. However, if the ultimate is to design an early warning system, it seems more natural to confront the models on the basis of their forecast ability.[2] Nonetheless, forecasting issues have only received superficial attention.

Some studies generate in-sample forecasts to compare rival models (Hajivassiliou, 1987; Detragiache and Spilimbergo, 2001). A few studies conduct out-of-sample evaluation but limited to a 1- or 2-year holdout period and using a fixed estimation window. The typical forecast criteria used in these studies are the Type I, Type II error or the overall error (Feder et al., 1981; Sommerville and Taffler, 1995; Oka, 2003; Peter, 2002).[3] Recent

[2]The issue of whether homogeneous or heterogeneous (linear) models provide better forecasts has been examined in the context of US gasoline and cigarette demand (Baltagi and Griffin, 1997; Baltagi et al., 2000).

[3]In this literature, the Type I (Type II) error is the missed defaults aand the Type II error is the false alarms. The Type I (Type II) error rate is estimated as the missed defaults (non-defaults) over the realized defaults (non-defaults). The overall error is the sum of missed defaults and false alarms over the number of sample cases.

contributions in the forecasting literature stress that such metrics may not represent well the decision-maker's problem (Granger and Pesaran, 2000; Pesaran and Skouras, 2002). Futhermore, those studies that provide out-of-sample predictions do not confront them to naive benchmarks. This is important due to the persistence of the external debt repayment behavior. A specific question we address is whether controlling for country, regional or time heterogeneities leads to 'better' sovereign default models.[4] The literature has not yet provided systematic evidence on the importance of heterogeneity in this context. This chapter seeks to fill this gap.

This chapter contributes to the literature in three directions. First, regarding model specification we consider a wide range of panel logits, some of them novel in this context, that treat regional-, country- and time-specific effects in different ways. The analysis is based on data for 96 countries over 1983-2002. The regressor set includes three world variables — macroeconomic uncertainty, monetary policy uncertainty and risk aversion — that, to our knowledge, have not been considered in the literature. Second, the models are compared using various statistical metrics that gauge their ability to describe the data generating process (DGP). These include likelihood ratio (LR) or Hausman type tests and the AIC and BIC model selection criteria. These metrics overwhelmingly suggest that heterogeneity across countries, regions and time should not be overlooked.

Third, a rigorous forecast analysis is conducted. A 12-year estimation window is rolled forward to generate recursive out-of-sample forecasts over a 5-year holdout period. We focus on 1-step-ahead point forecasts. Rather than using a fixed ad hoc cut-off probability, this parameter is optimally calibrated in-sample over each window for the model and loss function at hand. Both statistical and economic loss functions are evaluated over the holdout set and a positive-directional-change subset. The latter allows the emphasis to be on

[4]Forecast combining may be particularly fruitful when there is much uncertainty in finding the best model (Zou and Yang, 2004). This issue is investigated empirically for sovereign default in Chapter 3 of the thesis.

anticipating new (rather than perpetuating) debt crises. The equal-predictive-ability test of Diebold and Mariano (1995) is deployed to compare the models. Simple naive forecasts are also considered. These include random walk predictions and both the random prediction and the most-frequent-event prediction implicit in Pesaran-Timmermann's (1992) and Donkers-Melenberg's (2002) tests, respectively. Models that simply control for fixed regional- or time-effects are capable of yielding relatively good forecasts.

The chapter is structured as follows. Section 2.2 overviews the literature on country default risk, heterogeneity in debt repayment and evaluation of event forecasts. Section 2.3 describes the data and the endogenous default indicator. Section 2.4 outlines the models and the inference-based metrics. Section 2.5 discusses the forecast framework and Section 2.6 analyses the results. A final section concludes.

2.2 Literature Review

2.2.1 Empirical Evidence on Country Default Risk

A substantial body of literature has emerged, since the beginning of the Latin American debt crisis in 1982, on the empirical modelling of external debt crises and the prediction of default probabilities.[5] The existing can be grouped as follows. First, there are studies that take a totally descriptive approach aiming solely at uncovering the risk factors driving sovereign default on the basis of two theoretical approaches, the 'ability to pay' and the 'willingness to pay'. The former scheme relates to the idea of credit rationing in which the occurrence of default is a demand-supply disequilibrium situation where the international credit market does not clear at the interest rate ceiling.[6] The determinants of default are

[5] An excellent theoretical discussion on sovereign risk can be found in Obstfeld and Rogoff (1996), Cohen (1991) and Eaton et al. (1986). Comprehensive surveys of the empirical literature are provided in Cline (1995), Babbel (1996) and Aylward and Thorne (1998).

[6] In this case, the demand for new loans exceeds the maximum supply at the upper-ceiling interest rate at which bankers are willing to lend. As long as this demand-supply gap is less than the limit of arrears the lenders are willing to tolerate, the country may incur arrears on its debt-service obligations

those that shift the demand and supply curves to the right and to the left, respectively. For example, Edwards (1984), Callier (1985), Sommerville and Taffler (1995) and Manasse et al. (2003) focus on the macro- financial variables, Berg and Sachs (1988) on structural variables, Hajivassiliou (1989) on debt overhang.

According to the second scheme, 'willingness to pay' approach, the sovereign default event is the outcome of a utility cost-benefit comparison by a sovereign debtor (Eaton et al. 1986; Lee, 1991). Put differently, the borrower defaults on its external debt if the expected value of the discounted utility of consumption with default exceeds the expected value of discounted utility of consumption with repayment. The underlying theory postulates that even though the borrower possesses the financial resources to service debt, he may decide to hold back a debt-service payment that is due, if the benefits of default exceed the costs. The selection of explanatory variables for the analysis is based on this theory. Therefore, volatility of economic conditions and global financial links play the largest role. Solvency and liquidity variables are largely deemed irrelevant.

The second strand of the literature comprises various studies that produce default forecasts but that have mostly assessed their results on an in-sample basis. An indicative, but by no means exhaustive list, includes Saini and Bates (1984), Schmidt (1985), Callier (1985), Hajivassiliou (1987, 1994), Aylward and Thorne (1998), Detragiache and Spilimbergo (2002) and Catao and Sutton (2002). However, good in-sample properties do not ensure that the model can predict future debt crises, thus it is important that in sample analysis is supplemented by adequate out-of sample evaluation.

Finally, a small strand of the literature has forecasted default crises and documented out-of-sample evaluation of the results (Feder et al.,1981; Sommerville and Taffler, 1995; Peter, 2002; Oka 2003). These studies however employ a fixed estimation window, small validation samples and report Type I, Type II and overall error criteria for assessing

to cover the excess demand. If the excess demand rises beyond the arrears limit, a rescheduling of debt is perceived as a less costly option than the tolerance of further arrears for either borrowers or lenders.

forecast accuracy, which may not be very informative. A reason for the latter is that the stylized scoring rules (or loss functions) that may frequently not conform to the forecaster's decision problem. Several empirical studies highlight the marked *state dependence* of sovereign defaults (Aylward and Thorne, 1998; Brewer and Rivoli, 1997; Hajivassiliou, 1987, 1989, 1994; McFadden et al., 1985). Therefore, it is important to assess the predictive value of the model over and above the information implicit in the previous period's debt repayment status. To the best of our knowledge the literature does not compare the generated forecasts with naive benchmarks.

2.2.2 Heterogeneity in Debt Repayment

Earlier empirical work on sovereign default prediction was mostly based on logit or probit models that pool information across a variety of developing countries for a long time span. The need to allow for heterogeneity in the individual country parameters and over time was raised at an early stage by Saini and Bates (1984). The adequacy of this implicit homogeneity assumption in the country responses to fundamentals was later questioned by Solberg (1988), McFadden et al. (1985) and Hajivassiliou (1987, 1994) given the heterogeneity of the countries usually included in sovereign default analyses. Time-invariant country heterogeneity may be due differences in colonial history, religious or political factors and specific financial conditions, which altogether configure a country's attitude toward honouring external debt obligations and hence, lenders' credit decisions toward the borrowing country. Some of these country-specific effects are unobservable to the researcher and this poses difficulties for sovereign default modeling. For instance, the political or cultural environment within a country could be such that a homogeneous panel data model permanently underestimates its default probability.[7] In econometric terms, country heterogeneity could bias the parameters and cast doubt on their consis-

[7]An interesting example of cultural country-specific unobservables occurred after the price of crude oil fell dramatically in 1986. Many Islamic borrowers with significant indebtedness to US banks invoked the doctrine of 'sharia', which holds that the payment of interest is against the teachings of Koran.

tency. Aylward and Thorne (1998) point out that a possible reason why the majority of studies of external debt repayment behaviour, do not fully deal with the presence of unobserved country effects due to computational difficulties posed by panel logit models. Until recently, panel estimators were not available in the typical econometric packages.

The empirical evidence indirectly highlights the importance of the heterogeneity issue since very few macroeconomic variables have uniformly been found significant, while the significance of other factors varies from one study to another. The various contributions have restricted themselves to employing fixed or random effects specifications to capture time-invariant, country-specific unobservable effects. Within the framework of a random effects logit model, McFadden et al. (1985) find that 40 percent of the variation in countries' repayment performance is idiosyncratic or country-specific rather than attributable to homogeneous responses to the macroeconomic variables included in the model. They report statistically significant variability in the country-specific random effects term, however no evidence is provided on the possible gains in forecasting from a random error components specification. Hajivassiliou (1987) again finds statistically significant country-specific random effects and establishes a strong impact of such heterogeneity effects on in-sample crisis probabilities. Nonetheless, neither explicit measures of forecast accuracy nor out-of sample evaluations are reported. Subsequently, a few other studies including Li (1992) and Detragiache et al (2001) carry out sensitivity tests by employing fixed and random effects models to capture country heterogeneity and find that some of the variables lose significance with respect to the simple pooled logit. Oral et al. (1992) is the only study that allows for heterogeneity in the form of fixed country-specific slope parameters and intercept. Feder et al. (1981) and Solberg (1988), among others, use regional dummies to control for regional differences in the effects of the explanatory variables.

In the related context of currency crises, Kaminsky and Reinhart (1998) address the question of regional differences in the leading indicators. They present statistics to show that there are significant discrepancies in the volatility of leading indicators before the crisis in Asia, in Latin America and in other countries.[8] Drawing upon this finding, Burkart and Coudert (2002) employ a Principal Component analysis to confirm these regional differences and propose an estimator based on regional discriminant analysis specifications for Asia and Latin America, respectively. The authors find that the latter performs better than a global model that pools data across regions. They find some variables, such as the deviation of real exchange rate from trend, common to both specifications. Furthermore, they conclude that the regional contagion indicator which was significant in the global model no longer appears significant in the regional models This might indicate homogeneity *within* each region in terms of common structural problems. The latter are captured by the variables that are found to be specific for each region, namely, reserves to imports for Latin America and real domestic credit growth rate, openness and short term debt to total debt for Asia. In order to produce forecasts they then select for each the model (global or regional) that produces the best balance between Type I and Type II errors. As expected, this 'combined' model performs slightly better than the global model. However, they do not establish results on the out-of-sample performance. In general, only a handful of papers in the financial crisis literature explicitly consider regional models, the common practise being to use regional dummy variables.

With regards to heterogeneity over time the extant literature has only considered time-specific unobservable effects that are common across countries. In the context of a one-way fixed effects model for country repayment performance to the IMF and to external creditors in general, Aylward and Thorne (1998) and Solberg (1988) found significant

[8] Empirical research has shown evidence of strong links between currency crises and sovereign defaults, although the causal pattern is not clear. Reinhart (2001) find that in emerging makets the probability of a currency crisis conditional on having defaulted in 69%, while the probability of a sovereign default conditional on having experienced a currency crisis is 46%. Moreover, the literature has pointed out similar origins (leading indicators) for the two types of crises.

time-specific effects. This highlights the importance of global shocks on a sovereign's creditworthiness. The latter can also be captured through country invariant global variables such as interest rates or economic growth, which given the short time-series dimension in this type of studies tend to pick up trends and act as time-dummies. To the best of our knowledge, none of the studies in the literature have investigated the performance of models that account explicitly for time-heterogeneity (i.e. by allowing for different marginal effects of the leading indicators over time). Regarding the issue of whether the determinants of financial crises change over time Schnatz (1999) has examined two samples 1970-1997 and 1985-1997. He reports no significant difference in the determinants for the entire and restricted periods. This result may be a signal that increased global integration has not altered the structural factors that drive financial crises.

Berg and Pattillo (1999) and, more recently, Oka (2003) stress the need of further exploring specification issues, namely country heterogeneity and temporal dependence, in the context of panel binary-choice for financial crises. Thus, the empirical literature points towards the increased importance of heterogeneity in such panels of worldwide emerging markets and LDC. Despite the numerous theoretical justifications, a rigorous comparison among the homogeneous and the heterogeneous estimators has not been carried out. So far, little is known about the value added of heterogeneous estimators in an out-of-sample forecasting context.

2.2.3 Evaluating Event Forecasts

Forecast accuracy is one of the essential ingredients for discriminating between competing models. As stressed by Diebold and Mariano (1995) in their influential study, predictive performance and model adequacy are inextricably linked – predictive failure implies model inadequacy. The assessment of the forecast accuracy of binary-choice models deviates to some extent from the conventional forecast evaluation framework. The binary nature of the outcomes raises a number of issues that one is not confronted with in the case of continuous forecasts. Various measures of forecast performance pertinent to binary outcomes are summarized in Cramer (1999), of which the most frequently reported is the *hit rate* of the model. The hit rate is defined as the proportion of observations correctly predicted by the model.

A number of other measures have been used to evaluate probability forecasts associated of events that actually occurred, often referred to as scoring rules. The most popular is Brier's (1950) quadratic probability score (QPS), which is based on quadratic forecast errors, and is a rough analogue to the mean square error as routinely applied to point forecasts.[9] More recently, Diebold and Lopez (1996) discuss its application in forecasting default probabilities and advocate its use in conjunction with the Diebold-Mariano test. However, the QPS may not be appropriate in all contents, for instance if the loss is asymmetric. Other specifications include the asymmetric, logarithmic cost-function, applied in Diebold and Rudebusch (1989). The latter considers the logarithm of the forecasted probabilities and penalizes large errors more heavily. Diebold and Lopez (1996) present an excellent survey of forecast evaluation and comparison methods that covers ordinal and probability forecasts.

A crucial question that figures prominently in the recent forecasting literature is whether the widely applied forecast accuracy measures can lend themselves to representation of the economic loss when the forecasts are put into practice. Leitch and Tanner (1991) first argue that forecast evaluation should be based on profits (losses) realized upon using the forecasts and not on general, statistical criteria. A similar rationale is put forward by Pesaran and Timmermann (1994, 1995). Granger and Pesaran (2000a, 2000b) formalize the idea by developing a decision-theoretic forecast evaluation framework. In this context, the loss-function incorporates parameters to reflect the specific costs to individual decision-makers. They demonstrate the approach in a "two state-two action

[9] Note that it is not exactly analogous to the mean square error as it compares the forecasted probability with whether or not the event actually occurred and not with the actual probability.

decision problem", i.e. the binary-choice setup. General evaluation measures are abandoned in favour of one that can be tailored to the risk-preferences of the decision-maker. For instance, in the sovereign default context, irrespective of whether the forecaster is a policymaker, a financial institution or a multilateral bank her loss function will most likely place more weight on mispredicted defaults rather than false alarms or on new entries to default rather than defaults that are carrying over from a previous period. Pesaran and Skouras (2000) provide a review of decision-based approaches for evaluating and comparing forecasts. Pesaran and Timmermann (1992), Diebold and Mariano (1995) and McCracken (2000) elaborate on theoretical results for comparison and evaluation of out-of-sample predictions, respectively.

A good practice in event/probability forecast evaluation is to compare the resulting forecasts with 'naive' predictors to assess the value added over and above obvious guesses. The importance of the latter in a binary framework is further motivated by the misleading conclusions that can originate from a high prediction (hit) rate in an empirical scenario where the distribution between the two possible outcomes is uneven. For instance, in a problem like ours where the default event is rare a high hit rate may result only because the model predicts well the non-default states. Notwithstanding the ample room for overestimated forecast accuracy in the context of binary-choice models, naive model comparisons have received scant attention in the sovereign default literature.

Various naive predictors have been proposed in the binary-choice literature. 'Which one is best' is ultimately an empirical question that depends on the relative frequency of the two events. For instance, one naive model could forecast 0 or 1 randomly according to a certain probability. This naive forecast is implicit in the predictive dependence test for directional change developed by Pesaran and Timmermann (1992). Another option could be to use as naive predictor one that always forecasts the most frequent outcome. Donkers and Melenberg (2002) have derived the limiting distribution of the relevant test statistic for this case. Detailed exposition for these naive predictors follows in Section 2.5.

2.3 The Dataset

The analysis is based on *World Bank* data for 96 emerging markets and LDCs from Africa, Asia, Eastern Europe, Latin America and Middle East over 1983-2002. Information on external debt, arrears and rescheduling to official and private creditors is obtained from the Global Development Finance database. Macroeconomic and financial time series are obtained from the World Development Indicators database (see Appendix 2.1).

The Early Warning of Default (EWD) indicator

There is no unanimous definition of sovereign default. The rating agencies definition reflects only default on rated sovereign bonds, which is a rare event. Few countries have defaulted on rated bond issues, however, many countries defaulted on their bank debt and trade credit obligations, especially during the 1980s and the early 1990s. By adopting a broader definition that encompasses bank debt and trade obligations we can can base our study on a wider set of countries over a longer period of time.[10] We categorize country i as being in default in year t (denoted $d_{it} = 1$) if any of these conditions are met: a) there is a large jump in total arrears (interest and principal repayments) relative to total external debt $\Delta A_{it}/D_{it} > \delta$ or b) the total amount of debt rescheduled exceeds the decrease, if any, in total arrears.[11]

[10] Moody's and S&P's sovereign default rates are based on a limited number of defaults on rated sovereign bonds that occurred between 1998-2003. The leading rating agencies also report corporate default rates as proxies for the sovereign default probabilities associated with each sovereign credit-rating. Peter (2002) shows that these proxies are too low on average relative to the estimated default probabilities from a panel logit model.

[11] Some studies rely on reschedulings (Lee, 1991), others focus on arrears (Sommerville and Taffler, 1995; Aylward and Thorne, 1998) and a third group use both (Detragiache and Spilimbergo, 2001; Peter, 2002). Credit rating agencies also typically consider both arrears and reschedulings. Following Peter (2002), we adopt $\Delta A_{it}/D_{it}$ rather than A_{it}/D_{it} so that, say, countries with large arrears but which are reducing their arrears stock relative to external debt are not classified as default. We set the threshold δ for $\Delta A_{it}/D_{it}$ at its 1983-2002 mean $\hat{\delta} = 2.26\%$.

The goal is to predict the probability of a debt crisis over a specific time window longer than a year, while updating the forecasts annually. We adopt a 3-year warning window and define[12]

$$y_{it} = \begin{cases} 1 & \text{if } d_{i,t+k} = 1 \text{ at any } k = 0, 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

following Peter (2002) and Oka (2003). A unit value for this forward-looking variable, called the Early Warning of Default (EWD) state, signifies that country i has defaulted at least once over $[t, t+2]$. The default frequency in our sample is about 30% (see Appendix 2.2). The number of defaults per year is quite close to those identified by S&P (2001) for rated/non-rated debt.

Country-specific and global variables

A number of macroeconomic variables have been identified as determinants of sovereign ratings (Cantor and Packer, 1996), sovereign defaults (Detragiache and Spilimbergo, 2001) and sovereign spreads (Kamin and Kleist, 1999). Building on these findings we consider 25 domestic signals from the World Bank categories: *i*) external credit exposure, *ii*) external economic activity and financial resources, *iii*) conditions of real/public/financial sector and *iv*) global financial links. To reduce the degree of skewness and kurtosis in the ratios and the number of outliers, these are transformed using $sign(x)\ell n(1 + |x|)$. Any remaining outlier in each default/non-default group is reduced by windsorizing the ratios as follows. A data point x_{it} is indexed by $c \in \{0, 1\}$ according to whether it pertains to a tranquil ($y_{it} = 0$) or default ($y_{it} = 1$) window. If x_{it}^c falls outside $\bar{x}^c \pm 4\hat{\sigma}^c$, it is replaced by the appropriate interval limit.

The 5×1 world regressor vector, $\mathbf{z}_t^0$, includes two typical variables — the 10-year US Treasury Bond yield and OECD GDP growth — and three global indicators (annualized) that have not been used in the sovereign default context as yet. One is a proxy for

[12]The estimated models thus can be cast as "early warning" devices. Chapter 4 discusses how to optimally choose the warning horizon.

macroeconomic uncertainty obtained as the conditional variance of the US monthly logarithmic real GDP.[13] For this purpose, an appropriate AR(1)-GARCH(1,1) model is fitted to the detrended (first differenced) real GDP since the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests indicate the presence of a unit root in the levels. Second, a measure of *monetary policy uncertainty* is analogously derived from the monthly yield spread between the 3-month Treasury Bill and the US Federal Funds Target rate. According to the ADF and PP tests the spread is stationary and so an appropriate AR(1)-GARCH(2,1) is fitted to the levels for this purpose. Arora and Cerisola (2000) find evidence that sovereign bond spreads are significantly related to such uncertainty measure.[14] Third, the level of *global risk aversion* is proxied by the Sharpe ratio — the monthly average high-yield spread divided by its standard deviation over the last 12 months — based on the Merryl Lynch 175 US Corporate High-Yield index and the 10-year US T-Bond yield. Fitzgerald and Krolzig (2003) find evidence that this ratio captures risk aversion which dampens the demand of emerging market assets. The latter, in turn, influences capital flows and translates into lower FX reserve levels.

Large models typically have poor statistical properties. In order to preserve degrees of freedom, we deploy a cross-validation (jacknife) approach which assesses the relative value of each regressor on the basis of the in-sample Type I error (see Appendix 2.3).[15] The retained domestic and global signals are denoted by $\mathbf{x}_{it}$ and $\mathbf{z}_t$, respectively. These are discussed in Section 2.6.

[13]The quarterly US real GDP was interpolated into monthly frequency on the basis of the monthly industrial production (indicator series) using the proportional Denton approach that belongs to a family of LS-based benchmarking methods (Baum, 2001).

[14]The plot of the monthly spread suggests that there is an upward pattern during 1994 (Mexican crisis) and in the second half of 1998 (Asian crisis). The order of the GARCH is selected on the basis of a Ljung-Box test on the squared residuals. Details available from the authors upon request.

[15]Instead one could base the jacknife on some other criteria (e.g. the overall error rate) which may, of course, lead to a different regressor set. For our purpose, however, the relevant aspect is to use the same regressor set for all models in order to make the comparison informative with regard to the treatment of unobserved heterogeneity.

2.4 Models and Estimation

Let the observed EWD indicator, y_{it}, be influenced by a set of exogenous factors as follows

$$y_{it}^{*} = \alpha + \mathbf{x}_{it}'\beta + \mathbf{z}_t'\gamma + \varepsilon_{it}, \ \varepsilon_{it} \sim iid(0, \sigma^2), \ \ i = 1, ..., N, \ \ t = 1, ..., T \qquad (2.1)$$

where y_{it}^{*} is the latent index such that $y_{it} = 1$ for $y_{it}^{*} > 0$ and $y_{it} = 0$ otherwise. The noise ε_{it} is assumed independently distributed from the k domestic regressors ($\mathbf{x}_{it}$) and the r world regressors ($\mathbf{z}_t$). We have $p_{it} \equiv \Pr(y_{it} = 1) = \Pr(y_{it}^{*} > 0)$ and assuming a standard logistic distribution for ε_{it} it follows that $p_{it} = G(\mathbf{x}_{it}, \mathbf{z}_t) = \frac{\exp(\alpha+\mathbf{x}_{it}'\beta+\mathbf{z}_t'\gamma)}{1+\exp(\alpha+\mathbf{x}_{it}'\beta+\mathbf{z}_t'\gamma)}$. The response probability is thus the logit function evaluated at a linear function of the variables.[16] This nonlinear relation between p_{it} and $(\mathbf{x}_{it}, \mathbf{z}_t)$ can be rewritten linearly in terms of the log-odds ratio as $\ln \frac{p_{it}}{1-p_{it}} = \alpha + \mathbf{x}_{it}'\beta + \mathbf{z}_t'\gamma$. Equation (2.1) is referred to as the baseline *pooled logit* model (PLOGIT) which assumes full country and time homogeneity in the response y_{it}^{*} to $(\mathbf{x}_{it}, z_t)$. The $(1 + k + r) \times 1$ coefficient vector $(\alpha, \beta', \gamma')'$ can be estimated by Maximum Likelihood (ML).

2.4.1 Country-specific Heterogeneity

The PLOGIT can be extended to allow for unobserved country-specific effects α_i that stay constant over time, e.g. some countries are more likely to default than others in every period. The *fixed effects* model (FE) treats α_i as fixed and so there is an unknown $(N+k+r)\times 1$ coefficient vector to be estimated $\phi^{FE} = (\boldsymbol{\alpha}', \beta', \gamma')'$ where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ are country-specific constants. The error components or *random effects* model (RE) treats

[16] Both a standard normal and a standard logistic variable have a zero mean but the latter has a variance of $\pi^2/3$. Because the two pdf's are very similar (the logit density has more mass in the tails), if one corrects for the difference in scaling, the probit and logit models typically yield similar results in applied work. The main competitor to logit for classification is discriminant analysis. The latter assumes that the country's characteristics are multivariate normally distributed with a different mean vector (but the same variance-covariance matrix) associated to the default and non-default events. Most studies have concluded that logit is superior to discriminant analysis mainly because this normality assumption for the regressors is unrealistic (see Kennedy, 2003; ch. 15).

α_i as independent random draws from the same distribution with mean α and variance σ_α^2. Formally, $\alpha_i = \alpha + \sigma_\alpha v_i$ where $v_i \sim iid(0,1)$ is independent of $(\mathbf{x}_{it}, \mathbf{z}_t)$. Alternatively, it can be formalized as Eq. (2.1) with the composite error $e_{it} = \alpha_i + \varepsilon_{it}$. The $(2+k+r)\times 1$ parameter vector to be estimated is $\phi^{RE} = (\alpha, \sigma_\alpha; \beta', \gamma')'$.

Dependence between α_i and $(\mathbf{x}_{it}, \mathbf{z}_t)$ does not render $\hat{\phi}^{FE}$ inconsistent. However, the FE logit is bedevilled by two problems. One is the incidental parameters problem — inconsistency of $\hat{\alpha}_i$ for $N \to \infty$ and finite T — is transmitted into the slopes. This problem does not appear in the linear model because the α_i are effectively removed by using data in country-mean deviations. To avoid this issue, Chamberlain's (1980) conditional maximum likelihood (CML) estimator of the FE logit integrates the α_i out of the joint density by conditioning on $\sum_t y_{it}$. But then $\hat{\alpha}_i$ cannot be computed nor, in turn, the forecasts $\hat{p}_{it}$. The second problem arises from the fact that the FE model (linear or nonlinear) is only identified through the 'within' dimension of the data. If country i has the same status (y_{it}) in every period because, say, it has never experienced default, it is discarded in estimation. This may induce sample selection bias.[17]

The *random coefficients* parameterization (RC) goes one step further by allowing for random country heterogeneity both in intercepts and slopes. We consider two variants. First, a model (denoted RC$^\beta$) where the coefficients of the domestic regressors are heterogeneous

$$y_{it}^* = \alpha_i + \mathbf{x}_{it}'\beta_i + \mathbf{z}_t'\gamma + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma^2),\ i = 1, ..., N,\ y = 1, ..., T \qquad (2.2)$$

so that $\delta_i = (\alpha_i, \beta_i')'$ is a random vector with mean $E(\delta_i) = (\alpha, \beta')'$ and diagonal covariance matrix $E(\widetilde{\delta}_i\widetilde{\delta}_i') = \Omega$ with $diag(\Omega) = \{\sigma_\alpha, \sigma_{\beta 1}, ..., \sigma_{\beta k}\}$. Thus we can write $\widetilde{\delta}_i \equiv \delta_i - \delta = \Gamma \mathbf{v}_i$ where δ is a $(k+1)\times 1$ vector of fixed means, Γ is a diagonal matrix

[17] In this chapter, we do not consider lagged dependent variables $(y_{i,t-1})$ as regressors. The incidental parameters problem becomes more severe in dynamic models. The need to integrate out the α_i, in turn, prompts the initial conditions problem (see Greene, 2003; Ch. 21). Modeling dynamic effects and initial conditions in binary choice models is more complex than in the linear model.

such that $\Gamma\Gamma' = \Omega$, and $\mathbf{v}_i$ contains $(k+1)$ unobservable latent random terms which are $iid(0,1)$ and independent of $(\mathbf{x}_{it}, \mathbf{z}_t)$. The $(2+2k+r)\times 1$ parameter vector to be estimated is $(\alpha, \sigma_\alpha; \boldsymbol{\beta}', \sigma_{\beta 1}, ..., \sigma_{\beta k}; \boldsymbol{\gamma}')$.

Second, we consider a RC$^\gamma$ model where the effect of the global signals $\mathbf{z}_t$ on the log-odds ratio is country heterogeneous — equation (2.1) with the random vector $\boldsymbol{\delta}_i = (\alpha_i, \boldsymbol{\gamma}_i')'$. The $(2+k+2r)\times 1$ parameter vector to be estimated is $(\alpha, \sigma_\alpha; \boldsymbol{\beta}'; \boldsymbol{\gamma}', \sigma_{\gamma 1}, \sigma_{\gamma 2}, \sigma_{\gamma 3})$. Neither the RE nor the RC models (in contrast with FE) rely on large T for consistency. The FE logit is estimated by (C)ML whereas the RE, RC$^\beta$ and RC$^\gamma$ are estimated by maximum simulated likelihood (MSL).[18]

2.4.2 Region-specific Heterogeneity

In order to control for region-specific heterogeneity, each country is allocated into one of four groups: I) Asia ($N_{\mathrm{I}} = 17$), II) Latin America ($N_{\mathrm{II}} = 26$), III) Africa ($N_{\mathrm{III}} = 36$), IV) East Europe/Middle East/North Africa ($N_{\mathrm{IV}} = 17$).[19] We consider two approaches. First, the PLOGIT equation

$$y^*_{it,j} = \alpha_j + \mathbf{x}'_{it,j}\boldsymbol{\beta}_j + \varepsilon_{it,j}, \;\; \varepsilon_{it,j} \sim iid(0, \sigma_j^2), \quad i = 1, ..., N_j, \quad t = 1, ..., T \qquad (2.3)$$

is fitted to regions $j = \mathrm{I},...,\mathrm{IV}$.[20] This *regional logit* (RLOGIT) with $4(1+k)\times 1$ parameter vector $(\alpha_j, \boldsymbol{\beta}_j)$ can be seen as treating the regional heterogeneity in intercept and slopes as fixed. Second, a *regional regressor-specific* (RSLOGIT) model with $4 + \sum_j k_j$ parameters

[18]There is no closed form for the log-likelihood of the RC model. MSL involves draws from the multivariate density of $\mathbf{v}_i$. Bhat (1999) suggests $R = 1000$ draws and shows that a smaller number of Halton draws, $H = R/10$, is equally effective and cheaper. Our MSL uses a standard normal and $H = 500$. For the RE model this is asymptotically equivalent to the Hermite quadrature approach for approximating the likelihood (for details, see Greene, 2003).

[19]There are not enough degrees of freedom for the LOGIT estimation of any of these 3 regions so we group them. They share: i) a similar structure of exports given their oil exporting nature, ii) having gained access to international bond markets between 1995-98.

[20]Alternatively, one could estimate $y^*_{it} = \sum_{j=I}^{IV} \alpha_j D_j + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \varepsilon_{it} \sim iid(0, \sigma^2)$ by pooling all countries and using regional dummies. This is however a more restrictive version of (2.3) where $\boldsymbol{\beta}_j = \boldsymbol{\beta}$.

is considered where a distinct regressor set, $k_j \leq k$, is allowed for j = I,...,IV to preserve degrees of freedom.

2.4.3 Time-specific Heterogeneity

Equation (2.1) controls for common time effects (e.g. oil price shocks) by including the global signals $\mathbf{z}_t$. We consider also a *fixed time effects* (FTE) model that uses period dummies instead

$$y^*_{it} = \alpha_t + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma^2),\ i = 1, ..., N,\ t = 1, ..., T \tag{2.4}$$

where the $(T+k) \times 1$ vector $\phi^{FTE} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ with $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_T)'$ is estimated by ML.

Alternatively, the data can be conceptualized as a sequence of cross-section relations

$$y^*_{it} = \alpha_t + \mathbf{x}'_{it}\boldsymbol{\beta}_t + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma_t^2),\ i = 1, ..., N \tag{2.5}$$

over time $t = 1, ..., T$. The elements of the $T(1+k)\times 1$ vector $(\boldsymbol{\alpha}', \boldsymbol{\beta}')'$ where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_T)'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_T)'$ are obtained sequentially by ML. This approach allows for time variation in the intercept and slopes. On this basis we define the counterpart of the Pesaran and Smith's (1995) mean group estimator. Let $\hat{\beta}_{jt}$ denote the slope estimate of regressor j at period t. A *mean cross section* (MCS) estimator is defined as $\bar{\beta}_j^{MCS} \equiv (1/T)\sum_{t=1}^{T} \hat{\beta}_{jt}$ with standard error $SE(\bar{\beta}_j^{MCS}) = \sqrt{\frac{SD(\hat{\beta}_{jt})^2}{T}}$ where SD denotes the sample standard deviation.[21]

We should note that the mean MCS estimator provides a measure of $\boldsymbol{\beta} \equiv E(\boldsymbol{\beta}_t)$ whereas the above pooled time-series estimators measure $\boldsymbol{\beta} \equiv E(\boldsymbol{\beta}_i)$. A consensus view is that cross-section data estimate long run relations (Pesaran and Smith, 1995; Kennedy, 2003)

[21]If the slopes are random (orthogonal to $\mathbf{x}_{it}$), $\hat{\boldsymbol{\beta}}_t \rightarrow \boldsymbol{\beta}_t$ as $N \rightarrow \infty$ and then $\bar{\beta}^{MCS}$ is consistent as $T \rightarrow \infty$.

2.4.4 Time-varying Country Heterogeneity

Next we relax the assumption that the random country effects (in intercepts and/or slopes) are time invariant. More specifically, the RC$^{\beta}$ and RC$^{\gamma}$ models are generalized by allowing the coefficients to be time-dependent according to an AR(1) mechanism. Thus we have the RC$^{\beta}$-AR model

$$y_{it}^{*} = \alpha_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}_{it} + \mathbf{z}_{t}'\boldsymbol{\gamma} + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma^2) \tag{2.6}$$

where $\alpha_{it} = \alpha + \sigma_\alpha v_{it}^{\alpha}$ with $v_{it}^{\alpha} = \rho_\alpha v_{i,t-1}^{\alpha} + e_{it}$, $e_{it} \sim iid(0,1)$ so that $E(\alpha_{it}) = \alpha$ and $V(\alpha_{it}) = \frac{\sigma_\alpha^2}{1-\rho_\alpha^2}$; likewise for $\boldsymbol{\beta}_{it}$. The RC$^{\gamma}$-AR is analogously formulated.

These RC-AR formulations allow for the effects of the regressors on the log-odds ratio to vary across countries and over time. The $(3 + 3k + r) \times 1$ parameter vector of the RC$^{\beta}$-AR logit and the $(3 + k + 3r) \times 1$ vector of the RC$^{\gamma}$-AR counterpart are estimated by MSL.

2.4.5 Inference-based Metrics for Model Selection

Several metrics are employed to compare the above models. First, we use the BIC and AIC which have been shown by Monte Carlo simulation to have good finite-sample properties for a range of panel models (Hsiao and Sun, 2000). A ranking is thus obtained based on $AIC = -MLL + s$ and $BIC = -MLL + 0.5s\ln(NT)$ where s is the number of unknown parameters.

Statistical tests are also deployed. A Hausman test compares the FE and RE using the statistic $H = \mathbf{q}'\{V(\mathbf{q})\}^{-1}\mathbf{q} \overset{a}{\sim} \chi^2_{(s)}$ where $\hat{\mathbf{q}} = (\hat{\boldsymbol{\theta}}^{FE} - \hat{\boldsymbol{\theta}}^{RE})$, $V(\hat{\mathbf{q}}) = V(\hat{\boldsymbol{\theta}}^{FE}) - V(\hat{\boldsymbol{\theta}}^{RE})$ and s is the dimension of $\boldsymbol{\theta}$. The null is $\hat{\mathbf{q}} = 0$ and a rejection suggests that there are fixed effects and so the RE model is inconsistent.[22] This test can confront any two models such

[22]Unlike in the linear case, the RE estimator is inconsistent (in the presence of fixed effects) even if α_i is orthogonal to $(\mathbf{x}_{it}, \mathbf{z}_t)$. This is because in discrete-choice models the ML estimator is generally inconsistent under misspecification such as in the presence of unmeasured heterogeneity, omitted variables (even if they are correlated with the included ones) and any form of heteroskedasticity (Yatchew and

that both are consistent under the null but only the less efficient is consistent under the alternative.

The PLOGIT, RE, RC and RC-AR models are nested. For instance, under $H_0 : \sigma_\alpha = 0$ the RE collapses to the PLOGIT and thus latent country heterogeneity can be tested by a LR statistic (a counterpart of Breusch and Pagan's LM statistic) which is $\chi^2_{(1)}$ distributed. Likewise, the restrictions $\sigma_{\beta_1} = ... = \sigma_{\beta_k} = 0$ reduce RC$^\beta$ to the RE. For $\rho_\alpha = \rho_{\beta_1} = ... = \rho_{\beta_k}$, the RC$^\beta$-AR collapses to the RC$^\beta$. The PLOGIT and FE are also nested ($H_0 : \alpha_i = \alpha$) and the LR statistic follows a $\chi^2_{(N-1)}$ for large T and finite N. Caution is required with the latter test for large N.

The significance of the time effects in the FTE model can be tested with a LR statistic that has a limit $\chi^2_{(T-1)}$ distribution under $H_0 : \alpha_t = \alpha$. We also test for $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}$ and for $H_0 : \alpha_t = \alpha, \boldsymbol{\beta}_t = \boldsymbol{\beta}$ in the MCS model (which reduce it to the FTE and the PLOGIT without globals, respectively) using that $MLL_{MCS} = \sum_t MLL_{CS_t}$. The poolability across regions, $H_0 : \alpha_j = \alpha, \boldsymbol{\beta}_j = \boldsymbol{\beta}$ for $j =$ I,...,IV, can be assessed by means of a LR statistic which is $\chi^2_{3(k+1)}$ distributed.

2.5 Forecast Framework

Although the panel is unbalanced we denote the sample period by $[1, T]$ for expositional simplicity. A static model was used in Section 2.4 to simplify the theoretical exposition. In our analysis the regressors are lagged one year for forecasting purposes.[23] The last two years are also lost because of the forward-looking nature of y_{it}. Thus, in effect, y_{it} is observed over 1984-2000. The models are initially estimated over the first 12-year window, denoted $[1, T^*]$, and y_{i,T^*+1} is forecasted. This window is then rolled forward. Out-of-sample predictions are thus constructed over a holdout period $[T^* + 1, T]$ that spans $m = T - T^* = 5$ years (1996-2000) for $N = 96$ countries. This facilitates a

Griliches, 1984).

[23]This also helps to mitigate endogeneity bias.

relatively large Nm validation set.

The probability forecasts from, say, the PLOGIT model are $\hat{p}_{i,\tau+1}$ such that $\ln \frac{\hat{p}_{i,\tau+1}}{1-\hat{p}_{i,\tau+1}} = \hat{y}^*_{i,\tau+1}$ and $\hat{y}^*_{i,\tau+1} \equiv \hat{\alpha}_\tau + \mathbf{x}'_{i\tau}\hat{\boldsymbol{\beta}}_\tau + \mathbf{z}'_\tau\hat{\boldsymbol{\gamma}}_\tau$ is obtained over $[\tau - T^* + 1, \tau]$ recursively for $\tau = T^*, T^* + 1, ..., T - 1$. To forecast on the basis of the MCS model we recursively compute $\bar{\alpha}_\tau = (1/T^*)\sum_{t=\tau-T^*+1}^{\tau} \hat{\alpha}_t$ and $\bar{\beta}_\tau = (1/T^*)\sum_{t=\tau-T^*+1}^{\tau} \hat{\beta}_t$ and then construct $\hat{y}^*_{i,\tau+1} = \bar{\alpha}_\tau + \mathbf{x}'_{i\tau}\bar{\boldsymbol{\beta}}_\tau$. The probability $\hat{p}_{i,\tau+1}$ is transformed into an event forecast ($\hat{y}_{i,\tau+1} = 0, 1$) using a cut-off probability λ_τ which is optimally chosen for each model (see Chapter 4).[24]

Several forecast metrics are evaluated both over the Nm points in the *holdout* sample and over a subset called *positive-directional-change* (PDC) sample.[25] The latter excludes year t for country i if $d_{i,t-1} = 1$ so as to focus on the models' ability to predict default entry. This is important since debt crises, in contrast with currency/banking crises, are persistent (see Appendix 2.2).

We adopt statistical and behavioral loss functions. The latter can tailor more closely the forecaster's decision-making problem.[26] Pairwise model comparisons are drawn and additionally, each model is confronted to simple benchmarks. One is a RW type *event* model based on the last observed outcome $\hat{y}^{RW}_{i,\tau+1} = d_{i,\tau}$. Another is a RW type *probability* model, $\hat{p}_{i,\tau+1} = p_{i,\tau}$, where $p_{i,\tau}$ is the prior probability of default (the unconditional frequency of 1s) over the τ rolling window, namely, $\hat{p}^{RW}_{i,\tau+1} = \frac{1}{T^*}\sum_{t=\tau-T^*+1}^{\tau} y_{it}$. Finally, we consider the implicit benchmarks in the Pesaran-Timmermann (1992) and the Donkers-Melenberg (2002) tests.

[24]Extant studies use $\lambda = 0.5$ or fix it in-sample at the default frequency or at the value that minimizes the Type I and II error sum. For each rolling window τ, we find the λ_τ that minimizes the chosen loss function.

[25]Due to missing data we have a heterogeneous N_t, $t = 1, ..., m$ (or equivalently, m_i for $i = 1, ..., N$) holdout panel set. The ith country forecast loss over the holdout window is computed first, $\bar{L}_i = \frac{1}{m_i}\sum_{t=1}^{m_i} L(y_{it}, \hat{y}_{it})$ and then the overall loss $\bar{L}$ is obtained by averaging the latter over countries.

[26]See Diebold and López (1996) and Diebold and Rudebusch (1989) for comprehensive surveys.

2.5.1 Statistical Loss Functions

The *misclassification rate* (MR) defines the overall loss as the frequency of incorrect predictions

$$MR = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} y_{it}\{1 - I(\hat{p}_{it} > \lambda_t)\} + \{1 - y_{it}\}I(\hat{p}_{it} > \lambda_t), \quad MR \in [0,1] \tag{2.7}$$

where $I(\cdot)$ is an indicator function and λ_t is an optimal cut-off probability.

Scoring rules do not require λ_t because the probability forecast $\hat{p}_{it}$ is directly used. One is the *quadratic probability score* (QPS) or the Brier score which, strictly speaking, is not the direct counterpart of the MSE because it does not compare the realized event y_{it} with $\hat{y}_{it}$ but with $\hat{p}_{it}$

$$QPS = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} 2(\hat{p}_{it} - y_{it})^2, \quad QPS \in [0,2] \tag{2.8}$$

Second, the *logarithmic probability score* (LPS) defines the overall loss as

$$LPS = -\frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} y_{it}\ln(\hat{p}_{it}) + (1 - y_{it})\ln(1 - \hat{p}_{it}), \quad LPS \in [0,\infty) \tag{2.9}$$

and so it differs from QPS in that large errors are more heavily penalized.

2.5.2 Economic Loss Functions

Let the following *pay-off matrix* summarise the decision-making problem at hand[27]

		Actual state	
		$y_{it} = 0$	$y_{it} = 1$
Decision	$\hat{y}_{it} = 0$	ϕ_t^0	θ_t^1
	$\hat{y}_{it} = 1$	θ_t^0	ϕ_t^1

where θ_t^1 is the economic loss of a missed default and so forth ($\theta_t^j > \phi_t^j, j = 1,2$). We build on Granger and Pesaran's (2000) framework but make two simplifying assumptions: *a*) the cost of a correct forecast is zero, $\phi_t^0 = \phi_t^1 = 0$ and *b*) the cost of an incorrect forecast

[27]Granger and Pesaran (2002) define the economic cost of a decision based on the forecast $\hat{p}_{it}$ as $C_{it}(\hat{p}_{it}) = \phi_t^1 y_{it} I(\hat{p}_{it} > \lambda_t) + \theta_t^0(1 - y_{it})I(\hat{p}_{it} > \lambda_t) + \theta_t^1 y_{it}(1 - I(\hat{p}_{it} > \lambda_t)) + \phi_t^0(1 - y_{it})(1 - I(\hat{p}_{it} > \lambda_t))$.

is constant over the holdout period, $\theta_t^1 = \theta^1$, $\theta_t^0 = \theta^0$. We define the latter in relative terms, i.e. $\theta = \frac{\theta^1}{\theta^1+\theta^0}$.

The following *economic missclassification rate* (EMR) measure

$$EMR_\theta = \frac{1}{Nm}\sum_i\sum_t \theta y_{it}\{1 - I(\hat{p}_{it} > \lambda_t)\} + (1-\theta)\{1 - y_{it}\}I(\hat{p}_{it} > \lambda_t) \tag{2.10}$$

provides a family of forecast criteria for $\theta \in [0,1]$, each giving the overall loss associated to the model predictions ($\hat{p}_{it}$) for a particular decision-making scenario (θ). The forecast ranking from $\text{EMR}_{0.5}$ amounts to that from (7) since $\text{EMR}_{0.5} \equiv \frac{1}{2}$MR. The MR identically penalises missed defaults and false alarms like the hit rate given by HR=1-MR=$\frac{1}{Nm}\sum_{t=1}^m \sum_{i=1}^{N} [y_{it} \times \hat{y}_{it} + (1 - y_{it}) \times (1 - \hat{y}_{it})]$. In practice, the Type I and Type II errors need not be symmetric in their relative importance. For instance, from investors' viewpoint misjudging a bad borrower implies a fall in assets (reflected in the balance sheet) whereas incorrectly dismissing a loan applicant as a bad risk just entails a missed profitable lending opportunity. Hence, the average cost of a Type I error is typically higher than that of the Type II error. Nevertheless, we consider $\theta \in \{0.8, 0.2\}$ in the analysis below.

2.5.3 Forecast Accuracy Tests

Let $e_{it} \equiv L(y_{it}, \hat{y}_{it}^A) - L(y_{it}, \hat{y}_{it}^B), i = 1, 2, ..., N, t = 1, 2, ..., T$ be the loss differential between models A and B. We deploy the Diebold-Mariano (1995) [DM] test statistic

$$DM = \frac{\bar{e}}{\sqrt{\hat{f}/N}} \overset{a}{\sim} N(0,1) \tag{2.11}$$

where $\bar{e} = \frac{1}{N}\sum_i \bar{e}_i$ and $\hat{f}/N$ is an estimate of the variance of $\bar{e}$ that accounts for time dependence.[28] This test can readily accommodate non-normality of the forecast errors

[28] We have $\bar{e}_i = \frac{1}{m}\sum_t e_{it}$. The variance estimator is $\hat{f} \equiv V(\bar{e}_i) = \frac{1}{m^2}\sum_t V(e_{it}) + \frac{2}{m(m-1)}\sum_t\sum_{s>t} cov(e_{it}, e_{is})$ where $V(e_{it}) = \frac{1}{N-1}\sum_i(e_{it} - \bar{e}_t)$ for $t = 1, ..., m$ and $cov(e_{it}, e_{is}) = \frac{1}{N-1}\sum_i(e_{it} - \bar{e}_t)(e_{is} - \bar{e}_s)$. We also deployed the test by computing $DM_t = \frac{\bar{e}_t}{\sqrt{\hat{g}_t/N}}$ where $\bar{e}_t = \frac{1}{N}\sum_{i=1}^N e_{it}$

and is applicable to a wide class of loss functions (Diebold and López, 1996).[29] Below we deploy (11) also to confront each model with the appropriate RW type naive model, y^{RW} or p^{RW}, depending on the loss function.

Pesaran and Timmermann (1992) [PT] propose a nonparametric approach to test the null that forecasts and realizations are independent. The test is formulated on the basis of the HR. The total number of correct out-of-sample predictions (Nm times the model's HR), can be treated as a binomial random variable (the number of successes in Nm trials) with mean $Nm\tilde{p}$ and variance $Nm\tilde{p}(1-\tilde{p})$ where $\tilde{p} = \Pr(\hat{y}_{it} = 1, y_{it} = 1) + \Pr(\hat{y}_{it} = 0, y_{it} = 0)$. Under the null, $\tilde{p} = \hat{P}P + (1-\hat{P})(1-P)$ where $P \equiv \Pr(y_{it} = 1) = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} y_{it}$ and $\hat{P} \equiv \Pr(\hat{y}_{it} = 1) = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m} \hat{y}_{it}$ are the unconditional probability of observed and forecasted EWD states, respectively. Testing for independence amounts to comparing the model's HR with that of the implicit benchmark that predicts 1 randomly with probability $\hat{P}$. Thus we have

$$PT = \frac{HR - HR^{PT}}{\sqrt{(Nm)^{-1}\tilde{p}(1-\tilde{p})}} \overset{a}{\sim} N(0,1) \tag{2.12}$$

where $HR^{PT} \equiv \tilde{p}$ and a significant PT statistic suggests that the forecasts are dependent on the quantities to be predicted. However, as argued by Donkers and Melenberg (2002), predictive dependence does not imply that the model outperforms an uninformative naive model whose out-of-sample forecast is the in-sample outcome that is most often observed.

Donkers-Melenberg's (2002) [DoM] test for $H_0 : HR = HR^{DoM}$ is based on a naive model that predicts 0 in our setting. Thus we have $HR^{DoM} = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}(1 - y_{it})$. It

and $\hat{g}_t = V(e_{it})$. If dependence between DM_t and DM_s is assumed, then $DM = \frac{1}{m}\sum_{t=1}^{m} DM_t \sim N(0, \frac{1}{m})$. The results from the latter are qualitatively similar to those reported from (2.11) but the statistics are slightly higher. Finally, we considered the test variant $DM = \frac{\bar{e}}{\sqrt{\hat{w}/m}}$ where $\bar{e} = \frac{1}{m}\sum_t \bar{e}_t$ and $\hat{w} \equiv V(\bar{e}_t) = \sum_{k=-w}^{w} C_k(\bar{e}_t)$ for truncation lag $w = m^{1/3}$. Unsurprisingly, this long-run variance is very small ($C_k, k \geq 0$ is computed over just $m \leq 5$ points) and the resulting DM statistics are very large.

[29]The $N(0,1)$ density applies if the models are non-nested (McCracken, 2000). This has been shown by simulation to hold for ordinal and probability forecasts (López, 2001). In most of our tests, the models are non-nested.

follows that

$$D \equiv HR - HR^{DoM} = \frac{1}{Nm}\sum_{t=1}^{M}\sum_{i=1}^{N}(2y_{it} - 1) \times \hat{y}_{it}$$

and $\sqrt{Nm}D$ follows a limit normal distribution with zero mean and $E\{(2y_{it} - 1)^2 \times \hat{y}_{it}^2\}$ variance under H_0. For binary variables, the latter equals $E(\hat{y}_{it}) = \Pr(\hat{y}_{it} = 1)$. The DoM statistic is

$$DoM = \frac{HR - HR^{DoM}}{\sqrt{\frac{1}{Nm}[\frac{1}{Nm}\sum_{i=1}^{N}\sum_{t=1}^{m}\hat{y}_{it}]}} \overset{a}{\sim} N(0,1)$$

and it can be shown that, when a model has positive predictive performance (i.e. it outperforms the DoM naive), the quantities to be predicted and the predictions are dependent while the opposite is not necessarily true. In this regard, the DoM test is more challenging than the PT test.

2.6 Results and Discussion

The results of the jacknife are reported in Table 2.1.[30] The first column reports the results based on the world PLOGIT starting from the initial 30×1 regressor vector $(\mathbf{x}_{it}^0, \mathbf{z}_t^0)'$. The retained regressor set contains $k = 13$ domestic signals $\mathbf{x}_{it}$ and $r = 3$ global signals $\mathbf{z}_t$ (see Appendix 2.4 for definitions). A number of regressors between 10 and 15 is the norm in the literature (see Peter, 2002). Nearly all the debt (*external credit exposure*) ratios have good predictive power, the exception being the short-term debt/reserves ratio. The remaining $\mathbf{x}_{it}$ indicate *external economic activity* (2 out of 5 variables retained), *domestic conditions* (5/10) and *global links* (1/4). Interestingly, the retained $\mathbf{z}_t$ are the US macroeconomic uncertainty, monetary policy uncertainty and risk aversion proxies.

Columns 2-4 report the means per state and a t-statistic to assess whether it has good discriminatory power. The regressors retained by the jacknife have a significant mean differential, except for the trade balance/GDP, the real exchange rate misalignment

[30] The empirical analysis is conducted using LIMDEP 8.

(RER) and the three global indicators. In most cases, the sign of the mean differential can be explained theoretically, e.g. the debt/GDP during pending crisis episodes ($y_{it} = 1$) is about twice its level during tranquil periods. The exception is the short term/total debt ratio (liquidity) which is counterintuitive (see Appendix 2.5).

Columns 5-8 denote the region-specific regressor sets that are obtained next by applying the jacknife to each of the four RLOGIT equations starting from the above 16×1 variables.[31] Interestingly, the variables that are dropped in all the regional models are those that are unable to discriminate (t-statistic) between the two states, the only exception being GDP growth. Eight indicators are retained both in the world model and in at least two regional models: five debt burden and liquidity indicators (external debt/GDP, official/total debt, short-term debt/total debt, IMF credit/exports), a measure of macroeconomic control (per capita GNP), a macroeconomic stability signal (volatility of p.c. GNP growth) and a measure of openness (total trade/GDP).

Three domestic indicators are deemed weak in terms of predicting default: trade balance/GDP, GDP growth and the RER. The per capita GNP emerges as a strong signal, in contrast with GDP growth, perhaps because it reflects wealth. Total trade/GDP, which measures the country's degree of trade openness, is identified as a good default predictor in contrast with trade balance/GDP which measures the country's competitiveness. The latter is reflected in the current account of the balance of payments and is closely linked to the RER. This finding supports the 'willingness-to-pay' (as opposed to 'ability-to-pay') theory according to which the opportunity cost of not servicing debt is relatively high for integrated economies.

[31] One should ideally apply the jacknife to each RLOGIT equation starting from the initial 30 regressors but this is unfeasible in terms of degrees of freedom. In order to preserve the latter in the regional models, we use a stricter jacknife where a variable is dropped if, in so doing, the Type I error does not increase by more than 1%.

2.6.1 Inference-based Comparison

Below we compare rival logit models using several statistical metrics evaluated over 1984-2000.

Model ranking by information criteria

The AIC can be cast as a discrepancy measure between the true model and a candidate whereas the BIC approximates the posterior odds probabilities in a Bayesian framework. In the context of nested models the latter can be interpreted as adjusting the size of a LR test with the sample size. Table 2.2 reports the MLL, the number of estimated parameters and the AIC and BIC statistics. The best model according to the BIC is the RC^{β}(ng) that allows for random country-specific effects but no time effects (*ng* stands for 'no global variables'). The FE(ng) is generally not favoured (with or without globals) by the BIC but is ranked first by the AIC which less heavily penalises for the large number of estimated parameters. Nevertheless, the fact that the FE model is estimated over a distinct, reduced sample $\tilde{N} = 53$ — the countries for which there is no variation in y_{it} are thrown out — calls for caution in comparing its MLL with that of the other models.

The second BIC-best model is the RC^{β} that allows not only for random country effects but also time effects by including global regressors. The AIC ranks the RC^{β}(ng) and RC^{β} as second and fourth, respectively. The RC^{β}-AR and RC^{β}-AR models where the country-heterogeneneous coefficients are allowed to change over time fare better than both PLOGIT variants and that FTE.

At the bottom of the ranking are the models that control for either time effects (PLOGIT, FTE, MCS) or regional effects (RLOGIT, RSLOGIT) but not for country effects. The RLOGIT is preferred over the PLOGIT vindicating the importance of the regional effects. The worst model is the MCS followed by the RSLOGIT.

Importance of country-specific effects

Table 2.3 presents the test results for hypotheses regarding the country-specific effects. The LR statistic for $H_0 : \alpha_i = \alpha$ using the same reduced $\tilde{N}$ sample for the unrestricted (FE) and restricted (PLOGIT) model is clearly significant.[32] The ML estimates of the FE are widely dispersed, $\hat{\alpha}_i$ ranging from 3.0 to 17.9 with a standard deviation $SD(\hat{\alpha}_i) = 3.4$. We calculate $Z(\hat{\alpha}_i) = \frac{\hat{\alpha}_i - \bar{\alpha}}{SD(\hat{\alpha}_i)}$ where $\bar{\alpha}$ is the sample mean of the $\hat{\alpha}_i$. A coefficient estimate is identified as an outlier in the distribution of $\hat{\alpha}_i$ if $|Z(\hat{\alpha}_i)| > 1$ given that $SD(\hat{\alpha}_i)$ is quite large. About 34% of the countries are outliers. Similarly, the FE(ng) model suggests 32% of outliers.

Moreover, the LR statistic for the homogeneity null ($H_0 : \sigma_\alpha = 0$) in the context of the RE model is also significant. The estimate $\hat{\sigma}_\alpha = 2.35$ (t-ratio=22.49) suggests a large dispersion.[33] The measure $\hat{\sigma}^2_\alpha/(\hat{\sigma}^2_\alpha + \sigma^2)$ where $\sigma^2 \equiv \pi^2/3$ indicates that 63% of the variation in debt-servicing performance that is unexplained by $(\mathbf{x}_{it}, \mathbf{z}_t)$ is due to unobserved (time-constant) country heterogeneity. The RE versus RC$^\beta$ comparison ($H_0 : \sigma_{\beta_1} = ... = \sigma_{\beta_k} = 0$) also suggests country heterogeneity in $\boldsymbol{\beta}_i$. Likewise, the RE versus RC$^\gamma$ test ($\sigma_{\gamma_1} = ... = \sigma_{\gamma 3} = 0$) indicates that the influence of the global factors on the log-odds ratio is country-specific also.[34,35]

Caution is needed in interpreting the latter set of tests because, for instance, under $\sigma_\alpha = 0$ the parameter is on the boundary of the maintained hypothesis, $\sigma \in R^+ \cup \{0\}$. In such settings, the usual limit distribution of the test may not apply. For the case of a single restriction, an easy correction has been suggested — to use the χ^2 critical value for percentile $1 - 2\alpha$ instead of $1 - \alpha$ where α is the nominal level (see Kodde-Palm, 1986). Most obviously, the corrected test for PLOGIT versus RE remains significant. For joint

[32]The brute force test for $H_0 : \alpha_i = \alpha$ gives $LR = 470.8(0.00)$. This result must be interpreted with caution because the MLL_{FE} is obtained from a reduced set ($\tilde{N} = 53$) whereas the MLL_{PLOGIT} is based on all N countries.

[33]For the RE(ng) the estimates are similar at $\hat{\sigma}_\alpha = 2.37$ (t-ratio=23.00).

[34]We also tried a RC formulation which allows for heterogeneity in the 17×1 parameter vector $(\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}')$ but the MSL estimates and standard errors were massive. Hence, we discard this model as implausible.

[35]The $\hat{\sigma}_\beta$ and $\hat{\sigma}_\gamma$ estimates in the RC$^\beta$ and RC$^\gamma$ models, respectively, suggest large country heterogeneity also.

hypotheses (e.g. PLOGIT versus RC^{β}) the correction gets more complicated and is not pursued here. Nevertheless, the test statistics are rather large and so the corrected values are likely to remain significant.

Next we compare the PLOGIT model (ML) estimator and the FE model (Chamberlain's CML) estimator using a Hausman statistic. Under $H_0 : \alpha_i = \alpha$ both are consistent (the CML estimator is inefficient because *i*) it does not use this restriction, *ii*) it is based on a reduced sample) whereas under the alternative the consistent estimator is CML. The Hausman statistic at 32.14 strongly rejects. The regressor set without globals gives a qualitatively similar result at $34.46(0.001, 13)$.[36] The Hausman test for FE versus RE is clearly insignificant and so the latter is preferred.[37] In the comparison between RC^{β} and RE, on the one hand, and between RC^{γ} and RE, on the other, the Hausman test selects the RC models.

Importance of the time-specific and region-specific effects

Next we focus on the statistical importance of the time effects which are controlled for in distinct ways. First, a PLOGIT that includes $\mathbf{z}_t$. Second, the FTE that includes year-specific dummies. Third, the MCS that allows for time variation in the intercept and slopes. Fourth, the RC^{β}-AR (and RC^{γ}-AR) model that extends the RC^{β} (and RC^{γ}) formulation to allow for time-variation in the random country-specific slopes. Table 2.4 reports the results. The variables $\mathbf{z}_t$ are clearly significant ($H_0 : \boldsymbol{\gamma} = \mathbf{0}$) in the PLOGIT. A regression of the FTE estimates $\hat{\alpha}_t$ against $\mathbf{z}_t$ indicates that about half of the variation in the former ($R^2 = 46\%$) reflects shocks due to global macroeconomic uncertainty, monetary policy uncertainty and risk aversion. However, the relatively small Hausman statistic for PLOGIT(ng) versus PLOGIT at 4.12 supports the former specification. The joint

[36] Likewise, the Hausman test based on the Huber-White estimate of the sampling covariance matrix to account for unspecified latent heterogeneity is $31.16(0.01, 16)$.

[37] The Hausman test statistic to compare the RE estimate with Chamberlain's FE estimate is very similar at 7.75 (0.96,16). Also, the Hausman test to compare the FE(ng) and RE(ng) gives $7.98(0.84, 13)$.

restriction $H_0 : \alpha_1 = ... = \alpha_{T-1} = \alpha$ on the year dummies in the FTE model is rejected at the 10% level and several of the individual test statistics ($H_0 : \alpha_t = \alpha$) are significant at the 1% level. The Hausman statistic for FTE versus PLOGIT(ng) is clearly insignificant supporting the latter specification.[38]

The LR statistic for $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}$ (or $H_0 : \alpha_t = \alpha, \boldsymbol{\beta}_t = \boldsymbol{\beta}$) in the MCS model is insignificant. But this outcome may be an artefact of the huge number of restrictions being tested (above 200).[39] Indeed, the individual slope estimates suggest that there is marked time heterogeneity. For instance, the coefficient on GDP growth and the volatility of GNP p.c. growth have ranges [-18.84, 18.16], [-29.21,27.89] and standard deviations of 10.12 and 13.55, respectively. For each of the $j = 1, ..., 13$ regressors, about 30% of $\hat{\beta}_{jt}, t = 1, ..., T$ are identified as outliers using $Z(\hat{\beta}_{jt})$.

The LR test for $H_0 : \boldsymbol{\rho}_\alpha = \boldsymbol{\rho}_\beta = 0$ in the RC$^\beta$-AR model is highly significant and the individual tests for $\boldsymbol{\rho}_\beta$ are significant in 11/13 cases. The null $H_0 : \boldsymbol{\rho}_\alpha = \boldsymbol{\rho}_\gamma = 0$ in the RC$^\gamma$-AR model cannot be rejected but 2/3 of the individual tests for $\boldsymbol{\rho}_\gamma$ are significant. These findings suggest that the influence of the domestic (and possibly the global) indicators $\mathbf{x}_{it}$ on the log-odds ratio of default varies both across countries and over time.

We now assess the importance of controlling for heterogeneity at a regional level in the intercept and slopes $(\alpha_j, \boldsymbol{\beta}_j)', j =$ I,...,IV. We use the same 13 regressors, $\mathbf{x}_{it}$, in each region and test for parameter stability across regions. The LR test indicates that the regional heterogeneity is significant even at the 1% level. This is borne out by the variation across the estimates of the four RLOGIT equations. For instance, the coefficient on debt/GDP, official/total debt and trade/GDP has a regional range of [5.20, 23.03], [4.19, 30.23] and [−12.60, −3.29], respectively. The dispersion of the FE estimates $\hat{\alpha}_i$ within regions is

[38] The counterpart Hausman statistics based on the 'sandwich estimator' (Huber-White) are 12.19 (0.51,13) for PLOGIT versus PLOGIT(ng) and 10.84 (0.62,13) for FTE versus PLOGIT(ng).

[39] In these sequential cross-section regressions, the asymptotic distribution of the LR test holds for fixed T and $N \rightarrow \infty$. As T gets large, the number of restrictions will get large as well and the LR test may not be appropriate.

$SD_{II} = 2.7$, $\text{SD}_{III} = 2.9$ and $\text{SD}_{IV} = 2.4$ (the FE is unfeasible for region I) which is lower than within all countries.[40] A Hausman test for homogeneity ($\alpha_i = \alpha$) within regions gives smaller statistics than the world panel at 18.67(0.00), 31.89(0.00) and 7.00(0.43) for j = II, III, IV, respectively.

Slope estimates: plausibility of the signs

The impact of the domestic indicators (on the probability of default) has a clearcut theoretical sign in 8 out of 13 cases.[41] This is denoted in parenthesis. Table 2.5 sets out the estimation results.[42] Three of the coefficients bear the correct sign and are significant in all models: external debt/GDP (+), official/total debt (+) and trade/GDP (-).[43] The credit to private sector/GDP is significantly negative in all models thus supporting the view that this ratio is a proxy for banking development which is linked with increased economic growth (Bekaert et al. 2002).[44]

The significantly negative effect of GNP per capita is correctly picked up by the FE, RC^{β} and RC^{γ} models. But only the RC^{γ}-AR captures the negative effect of GDP growth.

[40]Likewise, the world FE model gives $\hat{\alpha}_i$ with $SD_I = 1.2, SD_{II} = 3.1, SD_{III} = 3.1$ and $SD_{IV} = 1.5$.

[41]Eaton and Gersowitz (1981) and Lee (1991) set out a theoretical framework where the default probability hinges on the 'willingness-to-pay'. The higher the volatility of export growth (and of GNP), the more an exclusion from the international capital markets is feared and so the more willing it is to honour its debt (-). Peter (2002) advocates the 'ability-to pay' theory whereby volatile economies typically have large current account deficits (+). A weaker currency (positive RER deviation from trend) favours trade competitiveness and hence exports (-) but it means also a high debt burden in home currency and so, if debt is serviced mostly using GDP, the likelihood of default is higher (+); an overvalued currency implies a high risk of a currency crisis and hence of sovereign default (-).

[42]Let $y^* = \mathbf{x}'\beta + \varepsilon$, the marginal effect of x_j is $\partial p/\partial x_j = G(\beta'\mathbf{x})[1 - G(\beta'\mathbf{x})]\beta_j$. The sign of $\partial p/\partial x_j$ is that of β_j.

[43]An exception is *official/total debt* in the FE model (t-ratio=0.7). The *external debt/GDP* signals the ability to pay debt. A large *total trade/GDP* ratio signals openness and hence, the opportunity cost of default. Countries experiencing severe balance of payments problems are the most likely borrowers from official, multilateral institutions such as the IMF and so their *official/total debt* ratio is high.

[44]The FTE and PLOGIT slopes $\hat{\beta}$ are very close so the former are not reported to preserve space. The t-ratios of the PLOGIT(ng), PLOGIT, R(S)LOGIT based on the asymptotic covariance matrix adjusted for unspecified latent heterogeneity (White's robust 'sandwich estimator') are qualitatively similar. But the t-tests must be interpreted with caution regarding autocorrelation in the residuals (see Fuertes and Kalotychou, 2004b).

The positive effect of the short-term debt ratio is picked up by the PLOGIT(ng), PLOGIT, MCS and RC$^\beta$-AR (also in the PLOGIT(ng) for Africa and EastEurope/MidEast/NorthAfrica; see Appendix 2.6). For debt service/exports, only the RC$^\beta$ captures the expected positive sign although the coefficient is insignificant. For trade balance/GDP, the RC$^\beta$-AR model yields the expected negative signed coefficient although it is insignificant also. The AR estimates for RC$^\beta$-AR and RC$^\gamma$-AR are generally below 0.85 which is suggestive of stationary time-series dependence. Regarding the (unreported) coefficients of the global regressors, the US macro uncertainty has the expected (+) sign and is significant in all models (where included) except for the FE. The US monetary policy uncertainty has the correct (+) sign and is significant in the PLOGIT and RE models. The US risk aversion proxy has the correct (+) sign in the RC$^\beta$, RC$^\beta$-AR and RC$^\gamma$-AR. In contrast to the other models, the FE suggests that all three global indicators are insignificant.[45]

Although the individual MCS estimates show massive instability as noted above, their average is plausible and comparable in magnitude to the PLOGIT estimates (see Appendix 2.6). This instability is reflected in large standard errors and six insignificant MCS coefficients. Large time series variation is expected in developing economy models due to measurement error. The MCS regressions cannot allow for country heterogeneity and this provides a rationale for the instability of $\hat{\beta}_t$. If the omitted factors (idiosyncratic shocks) responsible for the country heterogeneity change over time, this would induce different biases in $\hat{\beta}_t, t = 1, ..., T$ which may, however, cancel out when averaged.[46]

[45]In the PLOGIT at least two global signals (macro and monetary policy uncertainty) are significant whereas in the RLOGIT equations none is significant. The latter may be an artefact of the smaller degrees of freedom.

[46]In a logit, if the true DGP contains x_{1t} and x_{2t} but $y^*_{it} = \beta_1 x_{1t} + \varepsilon_{it}$ is specified, then $\text{plim}\hat{\beta}_1 = \delta_1\beta_1 + \delta_2\beta_2$ where δ_1 and δ_2 are complicated functions of the unknown parameters. If there are individual differences, the CS regression is $y^*_i = \alpha + \mathbf{x}'_i\beta + e_i$ where $e_i = (\alpha_i - \alpha) + \mathbf{x}'_i(\beta_i - \beta) + \varepsilon_i = \varsigma_i + \varepsilon_i$ and ς_i represents the factors responsible for the heterogeneous responses whereas ε_i are true innovations.

2.6.2 Out-of-sample Forecast Comparison

In sum, the information criteria suggest that models that allow for country heterogeneity (and possibly time effects also) fare better than models that control for heterogeneity at a broader, regional level. The (R)LOGIT and models that exclusively control for time effects such as the PLOGIT with globals, FTE and MCS fare relatively worse. The LR and Hausman tests tend to suggest that unobserved effects (of time, country and regional type) should not be overlooked in modelling the probability of default. However, the model comparison on the basis of the coefficient estimates (signs) is mixed.

We now compare the models' ability to predict outside of the estimation sample. Table 2.6 presents the overall losses over the entire holdout sample and the PDC subset. The model ranking according to the percentage of Type I and Type II errors is similar to that on the basis of MR for $\theta = 0.8$ and $\theta = 0.2$, respectively. For each loss function, the minimum-forecast-error (in bold) model is contrasted with all other models using the DM test. Asterisks denote a significant loss differential. Interestingly, the RSLOGIT provides the best forecasts under the loss functions implicit in the QPS , LPS and $\text{EMR}_{0.8}$ criteria over the holdout sample and it significantly outperforms the RW-type naive benchmarks. The DM test suggests that the PLOGIT(ng), PLOGIT, MCS and FTE have similar forecast accuracy to the RSLOGIT for these loss functions. Under the MR loss, the R(S)LOGIT, MCS, FTE, RE(ng) and RC^{γ} forecast significantly better than any other model but not the naive. Under $\text{EMR}_{0.2}$, the PLOGIT (ng), PLOGIT, FTE, RE and RC^{γ} are the best models but again the naive generates equally satisfactory forecasts.

Over the PDC subset, the simple PLOGIT(ng) attains the minimum-loss according to the QPS and LPS criteria and it forecasts significantly better than the naive ($\hat{p}_{it}^{RW}$). Under QPS, the PLOGIT, RSLOGIT and FTE models forecast equally well whereas all other models forecast significantly worse. Under LPS, the PLOGIT(ng) and PLOGIT forecast significantly better than any other model. The PLOGIT, MCS, RSLOGIT and

RE models forecast better than any other model (including the naive $\hat{y}_{it}^{RW}$) for the MR loss. Under $EMR_{0.2}$, the best forecasts are obtained from the PLOGIT(ng), PLOGIT, FTE and $\hat{y}_{it}^{RW}$, all other forecasts are significantly worse. Finally, under $EMR_{0.8}$ the minimum-loss model is the RSLOGIT whereas the FE, RC^{β}, RC^{β}-AR, RC^{γ}-AR models and $\hat{y}_{it}^{RW}$ are clearly significantly worse even at the 1% level.

It is worth noting that under the $EMR_{0.2}$ loss, the best forecasting model cannot beat the RW naive. Given that our sample is dominated by 0s, the naive predictor $\hat{y}_{it}^{RW}$ is expected to perform well because this criteria unrealistically penalises more heavily the false alarms than the missed defaults. Nevertheless, the forecast accuracy of the PLOGIT(ng), PLOGIT and FTE is equivalent to that of $\hat{y}^{RW}$ over both the holdout and PDC sets.

We next turn to the predictor dependence (PT) and predictive performance (DoM) tests. The PT statistic is significant in all cases at better than the 1% level (except for the FE model) suggesting that there is positive dependence between predictions and realizations for all models. Unsurprisingly, the DoM test rejects less often since it is relatively more demanding. The hit rate of the naive predictor implicit in the DoM test ($HR^{DoM} = 0.59$) is significantly larger than that of the PT test (which varies with the model) but smaller than that of the RW naive ($HR^{\hat{y}^{RW}} = 0.80$) over the holdout sample.[47] Among the models which pass the DoM test, the FTE or RSLOGIT produce the largest statistics over the holdout sample and the PLOGIT gives the largest statistic over the PDC sample. The FE, RC^{γ}, RC^{γ}-AR, RC^{β} and RC^{γ}-AR models do not pass the DoM test.

In sum, the more general formulations such as FE, RE, RC^{γ}(-AR), RC^{β}(-AR) that allow for unobserved (random) heterogeneity across countries and possibly over time too are good descriptive models but do not predict well out of sample. In contrast, parsimonious

[47]The benchmark predictor implicit in the DoM test coincides with $\hat{y}^{RW}$ over the PDC set with a hit rate $HR^{DoM} = 0.66$. Also, $HR^{PT} = HR^{\hat{y}^{RW}} = 0.66$ over the PDC set whereas for all other models $HR^{PT} < 0.66$.

regional logit regressions (RSLOGIT) forecast relatively well and outperform the naive benchmarks. The simple PLOGIT, MCS and FTE models that exclusively control for time-specific effects also work quite well as early warning devices.

2.7 Conclusion

The empirical literature on sovereign default is quite vast but a systematic analysis of the importance of controlling for differences in behavior across countries and/or through time that are not captured by the included regressors is lacking. This chapter seeks to fill this gap. We estimate different panel logit variants ranging from a simple pooled regression to a general random coefficients model where each country has its own coefficients that are specific to each time period also. The relative quality of the models is gauged on the basis of statistical hypotheses tests, model selection criteria, theoretical judgements on the plausibility of the estimates and forecast metrics.

The LR and Hausman type tests point towards the more general formulations. According to the AIC and BIC, simple models that exclusively control for time or regional heterogeneity are ranked last. The comparison is unclear, however, in terms of the plausibility of the coefficient estimates. Four domestic indicators — external debt to GDP, official to total debt, total trade to GDP and credit to private sector over GDP — emerge as robust default signals together with global macroeconomic uncertainty.

The loss function affects the forecast ranking. The overall picture is that panel logits that exclusively control for either time or regional heterogeneity in a simple manner provide more accurate sovereign default estimates than models that allow for random (country and time) effects. Moreover, the out-of-sample forecast ability of the selected models is superior to the naive benchmarks for most of the loss functions, particularly when a heavier penalty is attached to missed defaults (than to false alarms) or when the emphasis is in predicting entry into a debt crisis period. Therefore it seems reasonable to conclude that the panel model that best describes the data does not necessarily generate accurate sovereign default forecasts. More complexity in the models means, in effect, adding extra terms in the forecast error variance. These findings may have implications for the appropriate assessment of sovereign default risk, a task which is currently being pursued by rating agencies and international investors.

TABLE 2.1
Regressor Set Selected for World and Regional PLOGIT Models

World panel				Regional panels			
		mean		(I)	(II)	(III)	(IV)
N=96	y=0	y=1	t-stat	N=17	N=26	N=36	N=17
A) Country-specific indicators							
Total external debt/ GDP	0.3879	0.7032	23.76*	✓	✓	✓	×
Official debt / Total debt	0.5633	0.5933	6.91*	✓	✓	✓	✓
Short term debt / Total debt	0.1244	0.1109	2.99*	✓	×	✓	✓
Debt service / Exports	0.1492	0.1843	6.70*	×	✓	✓	×
IMF credit / Exports	0.0828	0.1529	9.68*	✓	×	✓	✓
Volatility of export growth	0.1072	0.1391	6.14*	×	×	×	✓
Trade balance / GDP	-0.0797	-0.0822	0.40	×	×	×	×
Credit to private sector/ GDP	0.2746	0.1785	13.41*	×	×	×	✓
GDP growth	0.0398	0.0247	6.26*	×	×	×	×
GNP per capita	7.1236	6.4402	13.87*	✓	×	✓	✓
Volatil. of GNP p.c. growth	0.0435	0.0529	6.35*	✓	✓	×	×
Real exchange rate	0.1273	0.1110	0.74	×	×	×	×
Trade / GDP	0.5427	0.4737	6.93*	✓	✓	✓	✓
B) Global indicators							
US macroeconomic uncertainty	0.2246	0.2281	0.81	×	×	×	×
US monetary policy uncertainty	0.2566	0.2572	0.18	×	×	×	×
US risk aversion	1.0453	1.0458	0.03	×	×	×	×

The variable selection is conducted over the in-sample [1984, 1995] period using the jacknife.
(I) Asia, (II) Latin America, (III) Africa, (IV) East Europe/Middle East/North Africa.
t-stat is the statistic for the significance of the absolute mean differential.
*denotes significant at the 1% level.

TABLE 2.2
Model Comparison on the Basis of Information Criteria

Model type	Controlled effects	MLL	s	AIC statistic	AIC ranking	BIC statistic	BIC ranking
PLOGIT(ng)	—	-606.3	14	620.3	**12**	656.5	**10**
PLOGIT	time	-601.4	17	618.4	**11**	662.4	**12**
MCS	time	-491.8	238	729.8	**15**	1007.5	**15**
FTE	time	-594.4	30	624.4	**13**	702.0	**13**
RLOGIT	regional	-500.1	56	556.1	**10**	660.0	**11**
RSLOGIT	regional	-646.7	26	676.7	**14**	735.1	**14**
RE (ng)	country	-469.8	15	484.8	**8**	523.7	**3**
FE (ng)	country	-399.0	66	405.0	**1**	559.4	**7**
RC^{β} (ng)	country	-399.7	28	427.7	**2**	500.1	**1**
RE	country, time	-467.8	18	485.8	**9**	532.3	**4**
FE	country, time	-366.0	69	435.0	**3**	596.4	**9**
RC^{γ}	country, time	-457.8	21	478.8	**6**	533.1	**5**
RC^{γ}-AR	country, time	-455.7	25	480.7	**7**	545.4	**6**
RC^{β}	country, time	-408.6	31	439.6	**4**	519.8	**2**
RC^{β}-AR	country, time	-425.8	45	470.8	**5**	587.3	**8**

The criteria are $AIC = -MLL + s$ and $BIC = -MLL + 0.5s\ln(NT)$ where s is the number of estimated coefficients and NT(=1307) is the effective sample size.
$AIC = -\sum_{t=1}^{T} MLL_t + \sum_t s_t$ and $BIC = -\sum_t MLL_t + 0.5s_t \sum_t \ln(N_t)$ for the TCS model, where N_t is the no. of available observations per cross-section.
$AIC = -\sum_{j=1}^{R} MLL_j + \sum_j s_j$ and $BIC = -\sum_j MLL_j + 0.5s_j \sum_j \ln(NT_j)$ for the R(S)LOGIT, s_j are the number of coefficients, NT_j the data points per region (R=4).

TABLE 2.3
Statistical Significance of Country-specific Effects

Tests	Model type					
	FE	RE	RC^{β}		RC^{γ}	
A) Likelihood ratio						
null hypothesis	$\alpha_i=\alpha$	$\sigma_{\alpha}=0$	$\sigma_{\alpha}=0$, $\sigma_{\beta_1}=...=\sigma_{\beta_k}=0$	$\sigma_{\beta_1}=...=\sigma_{\beta_k}=0$	$\sigma_{\alpha}=0$, $\sigma_{\gamma_1}=...=\sigma_{\gamma_3}=0$	$\sigma_{\gamma_1}=...=\sigma_{\gamma_3}=0$
	PLOGIT	PLOGIT	PLOGIT	RE	PLOGIT	RE
test statistic	141.1***	267.3***	385.7***	118.4***	287.3***	20.0***
	(0.00, 52)	(0.00, 1)	(0.00, 14)	(0.00, 13)	(0.00, 4)	(0.00,3)
B) Hausman-type						
null hypothesis	CML=PLOGIT	RE=FE	RC^{β}=RE		RC^{γ}=RE	
test statistic	32.14***	7.559	292.2***		78.58***	
	(0.01, 16)	(0.96, 16)	(0.00, 16)		(0.00, 16)	

The p-values and degrees of freedom of the tests are reported in parenthesis. ML estimation based on the Newton method (PLOGIT, FE) or by MSL using Halton draws (RE, RP). CML denotes the Chamberlain's conditional ML estimator of the FE slopes.
The MLL_{FE} and MLL_{PLOGIT} for the LR test of $H_0 : \alpha_i=\alpha$ are based on the largest possible sample ($\tilde{N}$=53) for the former.
*, ** and *** denote significant at the 10% , 5% and 1% level.

TABLE 2.4

Statistical Significance of Time- and Regional-specific Effects

	Time effects						Regional effects
Tests	PLOGIT	FTE	MCS		RC^{β}-AR	RC^{γ}-AR	RLOGIT
A) Likelihood ratio							
null hypothesis	$\gamma=0$	$\alpha_t=\alpha$	$\beta_t=\beta$	$\alpha_t=\alpha, \beta_t=\beta$	$\rho_\alpha=0, \rho_\beta=0$	$\rho_\alpha=0, \rho_\gamma=0$	$\alpha_j=\alpha, \beta_j=\beta,$
	PLOGIT(ng)	PLOGIT(ng)	FTE (ng)	PLOGIT (ng)	RC^{β}	RC^{γ}	PLOGIT (ng)
test statistic	9.89**	23.82***	205.30	229.12	44.22***	4.03	215.58***
	(0.02, 3)	(0.09, 16)	(0.54, 208)	(0.39, 224)	(0.00, 14)	(0.40,4)	(0.00, 42)
B) Hausman-type							
null hypothesis	PLOGIT=PLOGIT(ng)	FTE=PLOGIT(ng)	—		RC^{β}-AR1=RC^{β}	RC^{γ}-AR1=RC^{γ}	—
test statistic	4.12	9.93	—		122.20***	1816.17	—
	(0.99,13)	(0.70,13)			(0.00, 16)	(0.00,16)	

The p-values and degrees of freedom of the tests are reported in parenthesis. Models without global variables are signified by ng.

*, ** and *** denote significance at the 10% , 5% and 1% level.

TABLE 2.5

Parameter Estimates of Default Probability Models 1984-2000

	latent effects											
		time		country	regional		country, time					
Variables	PLOGIT(ng)	PLOGIT	MCS	RE(ng)	RLOGIT	RSLOGIT	FE	RE	RC^{γ}	RC^{γ}-AR	RC^{β}	RC^{β}-AR
External debt/GDP (+)	7.86	8.18	11.81	9.43	10.40	9.19	10.05	9.95	9.83	15.99	15.25	16.20
	(13.8)	(13.9)	(6.8)	(21.9)			(6.6)	(22.4)	(20.8)	(17.7)	(14.2)	(15.5)
Offic debt/ Tot debt (+)	8.41	8.82	7.92	6.22	15.09	15.57	2.85	7.04	6.91	8.98	11.52	18.65
	(4.7)	(4.9)	(3.7)	(5.9)			(0.7)	(6.4)	(5.7)	(3.5)	(9.5)	(4.6)
ST debt/ Tot debt (+)	5.44	5.41	4.38	-1.31	10.65	12.27	-6.59	-1.02	-1.75	2.69	1.52	8.31
	(3.3)	(3.3)	(2.0)	(-1.2)			(-1.8)	(-0.9)	(-1.4)	(1.11)	(0.5)	(2.2)
Debt serv/Exports (+)	-3.42	-3.80	-6.14	-2.69	-5.03	-3.16	-3.61	-3.01	-2.47	-2.07	1.29	-5.72
	(-3.6)	(-3.9)	(-2.9)	(-4.1)			(-1.92)	(-4.4)	(-3.5)	(-1.7)	(1.0)	(-3.0)
IMF credit/Exports (+/-)	-1.92	-2.11	-2.44	-2.34	2.22	2.82	-4.00	-2.70	-3.17	-3.02	6.70	11.20
	(-2.6)	(-2.8)	(-1.24)	(-4.2)			(-2.1)	(-4.8)	(-4.9)	(-3.2)	(4.3)	(11.7)
Vol export growth (+/-)	2.23	1.97	1.09	3.31	-1.71	4.37	3.21	2.93	2.50	-1.55	-4.07	-4.88
	(2.6)	(2.3)	(0.6)	(6.7)			(2.4)	(5.6)	(4.3)	(-1.8)	(-3.3)	(-3.5)
Trade balance/GDP (-)	1.23	1.05	0.73	3.80	-0.36	—	3.02	3.68	4.4	9.46	5.26	-0.91
	(1.6)	(1.4)	(0.7)	(6.1)			(1.5)	(5.7)	(6.2)	(7.7)	(3.7)	(-0.5)
Credit private /GDP (+/-)	-3.24	-3.49	-4.67	-3.66	-2.44	-4.73	-1.52	-3.81	-3.19	-15.61	-12.71	-16.53
	(-5.1)	(-5.3)	(-5.0)	(-7.8)			(-1.0)	(-8.2)	(-6.2)	(-12.6)	(-9.5)	(-14.8)
GDP growth (-)	-1.55	-1.24	-2.50	0.12	3.03	—	2.25	0.48	1.09	-8.91	9.17	9.42
	(-1.0)	(-0.8)	(-1.0)	(0.1)			(0.9)	(0.4)	(0.85)	(-5.2)	(3.74)	(3.9)
GNP per capita (-)	0.49	0.54	0.83	0.18	-0.13	-0.97	-1.97	0.26	-0.36	1.31	-2.30	0.95
	(4.8)	(5.2)	(5.1)	(2.1)			(-1.6)	(3.0)	(-3.5)	(8.1)	(-10.3)	(3.8)
Vol GNP pc growth (+/-)	4.55	3.51	-0.57	4.07	2.34	0.42	1.38	3.22	1.61	6.71	-7.16	-15.09
	(1.8)	(1.4)	(-0.2)	(2.2)			(0.3)	(1.7)	(0.8)	(1.6)	(-1.65)	(-2.52)
Real exchange rate[c] (+/-)	0.03	0.01	-0.16	0.15	0.01	—	0.14	0.13	0.12	0.81	-2.02	3.64
	(0.2)	(0.1)	(-0.4)	(1.4)			(0.6)	(1.1)	(0.9)	(7.2)	(-5.9)	(11.7)
Trade/GDP (-)	-4.70	-4.86	-7.31	-6.03	-7.14	-4.52	-6.36	-5.90	-5.74	-16.12	-7.58	-17.79
	(-8.0)	(-8.1)	(-4.8)	(-13.2)			(-3.1)	(-12.8)	(-11.7)	(-14.6)	(-7.7)	(-13.8)

The expected sign according to economic theory is in parenthesis after the variable name. The t-ratio of each coefficient is in parenthesis except for the RLOGIT and RSLOGIT for which the average across regions is reported (see Appendix C for details).

TABLE 2.6
Out-of-sample Forecast Analysis

Model	Statistical loss					Economic loss	
	MR	QPS	LPS	PT	DoM	$EMR_{0.2}$	$EMR_{0.8}$
A: holdout sample							
naive	**0.2014**	0.4117*	1.1588**	5.00**	3.73**	**0.0643**	0.1372**
PLOGIT(ng)	0.2700*	0.3275	0.5130	4.41**	1.90*	0.0754	0.0860
PLOGIT	0.2562*	0.3289	0.5168	4.51**	2.20*	0.0712	0.0927
MCS	0.2450	0.3435	0.5631	4.67**	2.39**	0.0973*	0.0776
FTE	0.2354	0.3367	0.5353	4.53**	2.87**	0.0703	0.0971*
RLOGIT	0.2314	0.3275	0.5904**	4.79**	2.72**	0.1004*	0.0835
RSLOGIT	0.2320	**0.3060**	**0.5043**	4.68**	2.82*	0.0960*	**0.0760**
RE(ng)	0.2434	0.3520*	0.6267*	4.36**	2.77**	0.0938*	0.0948
RE	0.2514*	0.3530*	0.6335*	4.24**	2.59**	0.0897	0.0948
FE	0.3898**	0.5251**	0.8874**	1.72*	0.29	0.1725**	0.1344
RC^{γ}	0.2492	0.3629*	0.6672**	4.51**	1.37	0.0836	0.0943
RC^{γ}-AR	0.3068**	0.5239**	2.2647**	3.61**	1.46	0.1388**	0.1236**
RC^{β}	0.3220**	0.6231**	3.1384**	3.06**	1.37	0.1581**	0.1588**
RC^{β}-AR	0.3088**	0.5752**	2.6447**	3.25**	1.61	0.1456**	0.1594**
B: PDC sample							
naive	0.3354**	0.4418**	1.2197*	0.00	0.00	**0.0671**	0.2683**
PLOGIT(ng)	0.3292**	**0.3489**	**0.5469**	3.38**	0.07	0.0819	0.0970
PLOGIT	0.2400	0.3491	0.5508	3.79**	2.92**	0.0776	0.1023
MCS	**0.2294**	0.3713*	0.6153**	3.92**	1.69*	0.1055**	0.0960
FTE	0.3228**	0.3621	0.5788**	3.16**	0.16	0.0817	0.1047*
RLOGIT	0.2716**	0.3992**	0.7290**	3.71**	1.57	0.1193**	0.1037
RSLOGIT	0.2474	0.3685	0.5935*	4.26**	1.83*	0.1233**	**0.0887**
RE(ng)	0.2648**	0.3776*	0.6718**	2.83**	1.22	0.1064**	0.0909
RE	0.2492	0.3743*	0.6730**	3.76**	2.01*	0.0996*	0.0951
FE	0.3642**	0.6321**	1.0806**	0.77	-1.12	0.2091**	0.1691**
RC^{γ}	0.2788**	0.3916**	0.7158**	3.11**	0.84	0.0900*	0.0947
RC^{γ}-AR	0.3172**	0.5483**	2.1234**	2.89**	0.24	0.1523**	0.1280**
RC^{β}	0.3174**	0.5994**	2.9334**	3.12**	0.61	0.1782**	0.1225**
RC^{β}-AR	0.3096**	0.5774**	2.5995**	2.16*	0.30	0.1470**	0.1658**

Bold denotes the minimum loss. **, * under (E)MR, QPS or LPS indicates that the model's forecast ability is significantly worse than that of the minimum-loss model according to a Diebold-Mariano test at the 1% and 5% level, respectively. PT is the Pesaran-Timmerman test of predictor dependence and DoM is Donkers-Melenberg test of predictive performance. The naive prediction is $\hat{y}^{RW}$ under MR, $EMR_{0.2}$ and $EMR_{0.8}$. The naive predictor is $\hat{p}^{RW}$ under QPS and LPS.

CHAPTER 3: Inferences On the Determinants of Default and the Autocorrelation Problem

3.1 Introduction

In the broad financial crisis literature a number of studies have developed EWSs, particularly, for currency and banking crises. These studies differ in the domestic indicators that they consider and in the model specification, typically some logit/probit variant. However, a common denominator across these studies is that they attempt to predict the probability of a crisis over a window which is typically longer than the frequency with which the forecasts are updated. More specifically, the dependent variable is a binary indicator such that a 1 value at period t indicates that the country has experienced a crisis over the $[t, t+h]$ window.[1] In Chapter 2 we have defined our EWD indicator over the $[t, t+2]$ time window, which facilitates annual sovereign default forecasts within a time window of three years. In a similar spirit, in the currency crisis literature, Berg and Patillo (1999) and Kaminsky et al. (1998) adopt a two-year horizon in their monthly forecasting models whereas the monthly updated predictions in Bussiere and Fratzscher (2002) and Burkart and Coudert (2000) refer to 3-, 12- or 24-month windows. In the sovereign default literature, Peter (2002) and Oka (2003) adopt a three-year horizon in their EWS for annual data. There is also a parallel macroeconomic literature which seeks to forecast recessions over a multiple period horizon — Bernard and Gerlach (1996) and Estrella and Mishkin (1997). The length T of the time series used in these studies is typically long enough for autocorrelation issues to become relevant. The fact that the dependent variable is a multiple-period outcome (overlapping problem) induces moving average prediction errors. On the other hand, the fact that the typical macroeconomic and financial ratios employed in these studies are themselves persistent results in autocor-

[1]See Abiad (2003) for a comprehensive survey of Early Warning Systems.

related prediction errors — a sluggish external debt/GDP indicator that predicts a debt crisis at t will also predict a debt crisis at $t+1$ and so on. Hence, the usual logit/probit ML standard errors will be biased and tests based on them have incorrect size.

However, the typical t-statistics reported in these and other studies to analyse the determinants of financial crises are based on the conventional ML standard errors. Little effort has been made on evaluation of misspecification effects on inference. At most, the Huber-White 'sandwich' estimator is reported but this is designed to account for neglected latent heterogeneity. To the best of our knowledge, Berg and Coke (2004) is the only study in the financial crisis literature that directly addresses the autocorrelation problem. As Berg and Coke point out, a likely reason why most EWS analyses do not pay attention to the presence of autocorrelated errors is that the usual econometric packages cannot produce an h.a.c. estimator for panel logit/probit models. They build on the autocorrelation-robust estimator for panel binary choice models proposed by Estrella and Rodrigues (1998) which is associated with the Generalized Method of Moments (GMM) technique. Estrella and Rodrigues (1998) and Berg and Coke (2004) show via simulation that the latter produces accurate standard errors whereas the usual probit ML standard errors are substantially downward biased.

The purpose of this chapter is to investigate empirically the extent of the above problem in the context of sovereign default for different logit specifications. These range from a baseline pooled model to a random coefficients model. We conduct tests for serial correlation in the generalized residuals to preliminary gauge the extent of the problem. Next we compute for each specification both the usual QML standard errors and the GMM type standard errors following Newey-West (1987). The results suggest that the inferences based on the usual QML standard errors are overturned when using the h.a.c. standard errors in a number of cases. This phenomenon is generally more clear for the macroeconomic ratios that are more sluggish as measured by the sum of the AR(p) coefficient estimates.

The chapter is structured as follows. Section 3.2 describes data and the conventional estimator. Section 3.3 outlines the serial-correlation problem inherent in panel EWS models and develops an autocorrelation-adjusted estimator. Section 3.4 presents and discusses the empirical results. A final section concludes.

3.2 Data and Estimation

The analysis in this chapter is based on annual data for 96 emerging and developing countries over 1983-2002. The exogenous variables are the 13×1 country-specific regressor vector, $\mathbf{x}_{it}$, and the 3×1 vector of global macrovariables, $\mathbf{z}_{it}$, picked up by the jacknife approach applied in Chapter 2. The goal is to predict whether debt-servicing problems are likely to arise within a specific time frame. We define the endogenous Early Warning Default (EWD) indicator as

$$y_{it} = \begin{cases} 1 & \text{if } d_{i,t+k} = 1 \text{ at any } k = 0, 1, .., h-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

using $h = 3$ years. Thus this forward-looking variable takes a value of 1 ('crisis warning') if a debt crisis has occurred sometime in a three-year window, i.e. in $t, t+1$ or $t+2$.

Let the observed EWD indicator, y_{it}, be influenced by exogenous factors as follows

$$y^*_{it} = \mathbf{x}'_{i,t-1}\beta + \varepsilon_{it}, \ \varepsilon_{it} \sim iid(0, \sigma^2), \quad i = 1, ..., N, \quad t = 1, ..., T_i \tag{3.2}$$

where the $(k+1) \times 1$ vector $\mathbf{x}_{i,t-1}$ includes a constant and y^*_{it} is the latent index such that $y_{it} = 1$ for $y^*_{it} > 0$ and $y_{it} = 0$ otherwise. The noise ε_{it} is assumed independently distributed from each of the k macroeconomic ratios included in $\mathbf{x}_{i,t-1}$. We have $p_{it} \equiv \Pr(y_{it} = 1) = \Pr(y^*_{it} > 0)$ and assuming a logistic density for ε_{it} it follows that $p_{it} = G(\mathbf{x}_{i,t-t}) = \frac{\exp(\mathbf{x}'_{i,t-1}\beta)}{1+\exp(\mathbf{x}'_{i,t-1}\beta)}$. The response probability is thus the logit function evaluated at a linear function of the variables. This nonlinear relation between p_{it} and $\mathbf{x}_{it}$ can be rewritten linearly in terms of the log-odds ratio as $\ln \frac{p_{it}}{1-p_{it}} = \mathbf{x}'_{i,t-1}\beta$. The maximum likelihood (ML) estimates of $\boldsymbol{\beta}$ facilitate the default probability estimates p_{it}, $i = 1, ..., N$, $t = 1, ..., T_i$..[2]

The baseline pooled logit model (PLOGIT), formalized by equation (3.2), assumes full homogeneity — the effect of a change in $\mathbf{x}_{i,t-1}$ on the log-odds ratio is identical across countries and time periods, $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}$. Alternative panel logit specifications are considered. On the one hand, in order to control for unobserved time-specific heterogeneity we specify either a PLOGIT which includes the global macrovariables $\mathbf{z}_{t-1}$ or a 1-way fixed time effects regression (FTE). On the other hand, we control for country effects by estimating either a fixed effects (FE), a random effects (RE) or a random coefficients (RC) model which include the global variables z_{t-1} to account for country-invariant time effects. The RC is a generalization of the RE where the time invariant country-specific heterogeneity is captured not only in a random intercept term (α_i) but through random slopes also. The FE, RE and RC models assume that the presence of the time-invariant country effects (e.g. α_i) captures all correlation between unobservables in different time periods. That is, the idiosyncratic error ε_{it} is assumed to be uncorrelated over individuals and time.

The PLOGIT and F(T)E regressions are estimated by the usual maximum likelihood (ML) approach whereas the RE and RC logit variants are estimated by maximum simulated likelihood (MSL) using 500 Halton draws. The notation *ng* (no globals) indicates that the regressor set is $\mathbf{x}_{i,t-1}$, otherwise is is $(\mathbf{x}_{i,t-1}, \mathbf{z}_{t-1})'$, and the slope coefficient vector of $\mathbf{z}_{t-1}$ is denoted $\boldsymbol{\gamma}$. Thus, for instance, RC$^{\beta}$ denotes a logit that allows for random coefficients $\boldsymbol{\beta}_i$ and RC$^{\gamma}$-AR denotes a logit with random coefficients $\boldsymbol{\gamma}_{it}$ (including a constant) which are time-varying according to an AR(1) mechanism and so forth.

[2]This is sometimes referred to as quasi-ML (QML) in the logit context because it does not incorporate assumptions about the time series model followed by ε_{it}. The (Q)ML method uses the likelihood corresponding to independent observations, i.e. $V(\varepsilon) = \sigma^2\mathbf{I}., \sigma^2 = \frac{\pi^2}{3}$.

3.3 Serial Correlation Problem and Correction

An EWS with a warning horizon that exceeds the interval between observations, i.e. $h > 1$, implies a moving average (MA) process in the prediction errors by construction. Moreover, the macroeconomic ratios typically employed as regressors are positively autocorrelated and this induces further serial dependence in the errors. In the context of non-linear logit/probit models, time series dependence in the errors does not hinder the consistency of the coefficient estimates, provided that regressors and errors are uncorrelated (exogeneity).[3] However, the existence of serial autocorrelation will invalidate conventional t- and F-tests which use the standard ML covariance matrix estimator $\hat{\Omega}$. The latter yields biased standard errors.

One way to avoid misleading inferences, without the need to impose specific assumptions on the structure of the covariance matrix of the parameter estimators, Ω, is the use of the usual ML logit estimator while adjusting the standard errors for general forms of autocorrelation. In this spirit, Gourieroux et al. (1984) and Poirier and Ruud (1988) propose autocorrelation robust estimators of the panel probit covariance matrix and demonstrate their consistency. However these estimators are computationally cumbersome.[4]

This chapter follows the GMM methodology along the lines of Newey-West (1987) proposed by Estrella and Rodrigues (1998) in the probit context. We adapt the latter to the logit framework as follows. The log-likelihood function of the logit model is

$$\ln L = \sum_{i=1}^{N}\sum_{t=1}^{T}\{y_{it}\ln G(\beta'\mathbf{x}_{it}) + (1-y_{it})\ln[1-G(\beta'\mathbf{x}_{it})]\}$$

and the first-order (maximization) conditions are:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N}\sum_{t=1}^{T}[\frac{y_{it}g_{it}}{G_{it}} + (1-y_{it})\frac{-g_{it}}{(1-G_{it})}]\mathbf{x}_{it} = 0$$

[3] Wooldridge (1994) demonstrates the consistency of the probit/logit ML estimator under general conditions.

[4] For instance, they require numerically solving single and double integrals of the gaussian pdf.

where $G_{it} \equiv G(\beta' \mathbf{x}_{it}) = \Pr(y_{it} = 1)$ and $g_{it} \equiv g(\beta' \mathbf{x}_{it}) \equiv \frac{dG_{it}}{d(\beta' x_{it})} = \frac{e^{\beta' \mathbf{x}_{it}}}{(1+e^{\beta' \mathbf{x}_{it}})^2} = G_{it}(1-G_{it})$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - G_{it})\mathbf{x}_{it} = 0$$

and $(y_{it} - G_{it}) \equiv \hat{\varepsilon}_{it}$ is referred to as the generalized residual.[5] The ML estimator of the asymptotic variance-covariance matrix is

$$\hat{\Omega} = -\mathbf{H}^{-1}$$

where $\mathbf{H}$ denotes the Hessian matrix

$$\mathbf{H} = \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{N} \sum_{t=1}^{T} [G_{it}(1 - G_{it})]\mathbf{x}_{it}\mathbf{x}_{it}'$$

This estimator is inconsistent in the presence of serial correlation. The heteroskedasticity and autocorrelation consistent (h.a.c.) estimator associated with GMM techniques is as follows,

$$\hat{\Omega}^{hac} = \mathbf{H}^{-1} \hat{S}_* \mathbf{H}^{-1} \tag{3.3}$$

where

$$\hat{S}_* = \hat{S}_0 + \sum_{i=1}^{N} \left[\sum_{j=1}^{L} \sum_{t=j+1}^{T} \omega_j \tilde{\varepsilon}_{it} \tilde{\varepsilon}_{it-j} (\mathbf{x}_{it}\mathbf{x}_{it-j}' + \mathbf{x}_{it-j}\mathbf{x}_{it}') \right]$$

and $\hat{S}_0 = \sum_{i,t} \tilde{\varepsilon}_{it}^2 \mathbf{x}_{it}\mathbf{x}_{it}'$. The above expression provides a family of GMM type robust estimators for the asymptotic covariance matrix of the logit parameter estimator. For instance, one can use $\omega_j = 1$ in the spirit of Hansen's (1982) approach or alternatively, the Barlett window $\omega_j = 1 - \frac{j}{1+L}$ following Newey-West (1987) so that higher-order autocorrelations are given less importance; L represents the maximum plausible autocorrelation

[5]Note that the first derivative of the log-likelihood with respect to the constant term yields $\sum_i \sum_t (y_{it} - G_{it}) = 0$. This implication bears some similarity to the LS normal equations and so $y_{it} - G_{it}$ is viewed as a residual.

lag. Estrella and Rodrigues (1998) demonstrate that the $\hat{\Omega}^{hac}$ estimator in the probit case (i.e. where $\tilde{\varepsilon}_{it}$ are the ML generalized residuals of probit) has the same asymptotic properties as the estimators derived by Poirier and Ruud (1988) and Gourieroux et al. (1984). In particular, as long as L increases with the time series length T, the $\hat{\Omega}^{hac}$ estimator for the variance-covariance matrixis consistent. One advantage of (3.3) over the former estimators is that is easy to compute. Both Estrella and Rodrigues (1998) and Berg and Coke (2004) demonstrate by Monte Carlo simulation, although on the basis of different data generating processes, the satisfactory performance of $\hat{\Omega}^{hac}$ for small samples.

In our application, we adopt $L = 3$, given that $h = 3$ in our dependent variable definition above, and follow the Newey-West approach.[6] Our preliminary analysis of the data includes two residual autocorrelation tests — the Baltagi and Li (1995; BL) LM type test and the Bhargava, Franzini and Narendranathan (1982; BFN) Durbin-Watson type test. The alternative hypothesis is that there is first-order autocorrelation in the idiosyncratic error term ε_{it}. Random or fixed individual effects are allowed. However, like most existing panel autocorrelation tests, the BFN and BL tests have as pitfall that they assume homogeneity in serial correlation across individuals. The BFN statistic

$$BFN = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T_i} (\hat{\varepsilon}_{it} - \hat{\varepsilon}_{i,t-1})^2}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\varepsilon}_{it}^2}$$

converges to 2 under H_0, i.e. it suggests first-order (positive) autocorrelation if $BFN < 2$. The BL test statistic

$$BL = \frac{N\tilde{T}^2}{\tilde{T}-1} \left(\frac{\sum_{i=1}^{N} \sum_{t=2}^{T_i} \hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\varepsilon}_{it}^2} \right)^2$$

where $\tilde{T} = \frac{\sum_{i=1}^{N} T_i}{N}$, suggests first-order autocorrelation if BL exceeds the usual 5th or 10th quantile of the $\chi^2_{(1)}$ distribution. In our context, $\hat{\varepsilon}_{it}$ denotes generalized residuals.[7]

[6] A longer lag implies estimating the second term of $\hat{S}_*$ on the basis of a shorter sample $(T - L)$. Experimenting with $L = 4$ produced qualitatively similar results.

[7] The theoretical properties of the BFN and BL statistics (bounds and sampling distribution, respectively) have been derived under balanced panels, i.e. $T_i = T = \tilde{T}$. Hence, the test results for our unbalanced panel should be taken as indicative only.

To provide a measure of the degree of persistence in each macroeconomic ratio, the pooled series are used to estimate an AR(p) model with sufficiently long lag p so as to absorb the underlying time dependence. The autoregression includes country dummies to control for autocorrelation arising from unobserved time-invariant country effects, i.e. $x_{it} = \alpha_i + \rho_1 x_{i,t-1} + ... + \rho_p x_{i,t-p} + \varepsilon_{it}$. The model estimates facilitate the persistence measure $\tau = \sum_{j=1}^{p} \rho_j$.[8]

3.4 Results and Discussion

Table 3.1 sets out the coefficient estimates for each logit model and the usual t-statistics.[9] In parenthesis we denote for each coefficient the expected theoretical sign. We use (+/-) to indicate that contradicting theories can be considered. The t-ratios for the PLOGIT and FTE models (that do not control for country-specific heterogeneity) using Huber-White's robust 'sandwich' estimator that adjusts for unspecified latent heterogeneity are reported in square brackets.[10] Although the t-ratios change, the results are qualitatively similar. However, all t-tests in Table 3.1 must be interpreted with caution because of residual autocorrelation. To gauge the extent of the latter we applied the Bhargava, Franzini and Narendranathan's DW test and Baltagi and Li's LM test to the generalized residuals of each model.

The BL and BFN test statistics are strongly significant in all models. The former ranges from a minimum of 472.09 (RC$^{\beta}$-AR residuals) to a maximum of 867.86 (RC$^{\beta}$) and similarly, the strongest rejection with the BFN test occurs for the RC$^{\beta}$ model at 0.292

[8]Observations $t = 1, 2, ..., 5$ are lost for all $i = 1, ..., N$. Note also that since $N = 96$ whereas $T = 17$ (or smaller T_i due to missing data for some countries) there may be a downward bias in the fixed effects AR(p) estimates (Nickel, 1981). Nevertheless, in our context the persistence measure $\sum \rho_j$ is not analysed in absolute terms for each variable but instead compared across them.

[9]The Hausman test statistic to compare the slope estimates in the FE and RE models gives a p−value of 0.96 which clearly supports the efficient RE estimates. Given the latter and the fact that the FE is based on a different sample from all other models (countries for which there is no time variation in y_{it} do not enter the log-likelihood function) the FE estimates are not reported to preserve space.

[10]The Huber-White covariance matrix can be seen as formula (3.3) for $\omega_j = 0$,

whereas the highest statistic is that for the RC$^{\beta}$-AR model at 0.700. For comparative purposes, we deployed the tests also for the same models where the dependent variable y_{it} is defined as in (3.1) with $h = 1$, i.e. the forecast horizon matches the annual frequency of the data. The BL and BFN statistics are now notably smaller but still significant, as expected, due to the persistence in the macroeconomic ratios and in the default indicator $d_{i,t}$ itself. For instance, the BL statistic falls from 628.81 ($h = 3$) to 149.05 ($h = 1$) for the PLOGIT and from 635.01 to 152.30 for the PLOGIT(ng). Likewise, the BFN statistic increases from 0.490 to 1.185 for LOGIT and from 0.491 to 1.170 for PLOGIT(ng). These figures illustrate that even if there is no overlapping problem, the sluggishness of the macroeconomic ratios is an important source of serial dependence in the logit prediction errors.[11]

Table 3.2 reports the t-statistics based on the Newey-West type standard errors and the AR(p) persistence measure for each regressor. For the latter, a maximum lag order $p = 5$ years seems to absorb most of the serial dependence in the errors.[12] The h.a.c. standard errors are larger than the usual QML (or Huber-White adjusted) standard errors by a factor that ranges between 1.5 and 18 times. The coefficients lose significance or become marginally significant in a number of models and so one should be cautious about the inferences from earlier sovereign default studies. For instance, the coefficients of GNP per capita and IMF credit/exports lose significance in five and three specifications, respectively, when the h.a.c. standard errors are used. Not surprisingly, the degree of persistence of these variables is relatively large as compared with the remaining ratios.

[11] We also compared the usual QML t-statistics and the h.a.c. t-statistics in the counterpart models for y_{it} defined using $h = 1$ (no overlapping problem). Although the latter are smaller than the former, the difference is now smaller. This is in line with the fact that the serial dependence in prediction errors stems now from the sluggish regressors and the persistence inherent in d_{it} itself. For instance, for GNP p.c. in the LOGIT (ng) the usual t-ratio is 3.87 whereas the h.a.c. t-ratio is 3.01, for debt service/exports the usual t-ratio is -3.96 and the h.a.c. t-ratio is -3.39 and so forth.

[12] For instance, for the three most sluggish variables — GNP per capita, credit to private sector over GDP and IMF credit to exports — the estimated autocorrelation in the residuals is negligible at -0.004, -0.039 and -0.027, respectively.

The more marked changes occur in the RC logits where several variables lose significance.

3.5 Conclusion

Autocorrelatedness in the prediction errors is a pervasive feature in an early warning system (EWS) of sovereign default, or of a broader financial crisis, for two reasons. One is that the macroeconomic and financial ratios typically employed are persistent and this results in serially dependent prediction errors. Another is that, by definition, the goal of an EWS is to estimate the probability of a debt crisis over a period (window) that exceeds the frequency with which the forecasts are updated. This overlapping problem induces a moving average process in the prediction errors. A large number of studies use panel logit/probit classifiers estimated by maximum likelihood to investigate the determinants of default. To the best of our knowledge, none of them raises the issue of autocorrelated errors and their impact on the inferences. This chapter is a step towards filling this gap.

We estimate different panel logit variants ranging from a simple pooled regression to a random coefficients model where each country has its own coefficients that are specific to each time period. We utilize Estrella and Rodrigues's (1998) Newey-West type covariance matrix estimator which is associated with GMM techniques. Using two conventional serial correlation tests, the logit prediction errors are shown to be significantly autocorrelated even when the forecast horizon matches the frequency of the data so that there is no overlapping problem. The overall picture that emerges from this analysis is that the adjusted standard errors are substantially higher than the conventional ML logit standard errors particularly for the most sluggish macrovariables. As a result some of the variables lose significance or become marginally significant. Our results are in line with the simulations in Berg and Coke (2004) for a panel probit calibrated using a currency crisis dataset.

TABLE 3.1

Logit EWS Models for Sovereign Default, 1984-2000 (QML Standard Errors)

	latent effects								
		time		country	country, time				
Variables	PLOGIT(ng)	PLOGIT	FTE	RE(ng)	RE	RC^{γ}	RC^{γ}-AR	RC^{β}	RC^{β}-AR
External debt/GDP (+)	**7.86**	**8.18**	**8.39**	**9.43**	**9.95**	**9.83**	**15.99**	**15.25**	**16.20**
	(13.77) [12.75]	(13.88) [12.72]	(13.96) [12.84]	(21.98)	(22.37)	(20.76)	(17.67)	(14.22)	(15.5)
Offic debt/ Tot debt (+)	**8.41**	**8.82**	**8.50**	**6.22**	**7.04**	**6.91**	**8.98**	**11.52**	**18.65**
	(4.7) [5.45]	(4.86) [5.47]	(4.57) [5.02]	(5.87)	(6.44)	(5.68)	(3.48)	(9.51)	(4.6)
ST debt/ Tot debt (+)	**5.44**	**5.41**	**5.21**	-1.31	-1.02	-1.75	2.69	1.52	**8.31**
	(3.30) [3.80]	(3.25) [3.73]	(3.09) [3.47]	(-1.18)	(-0.91)	(-1.40)	(1.11)	(0.50)	(2.2)
Debt serv/Exports (+)	**-3.42**	**-3.80**	**-4.26**	**-2.69**	**-3.01**	**-2.47**	-2.07	1.29	**-5.72**
	(-3.63) [-3.59]	(-3.94) [-3.78]	(-4.29) [-4.04]	(-4.05)	(-4.44)	(-3.47)	(-1.7)	(0.98)	(-3.0)
IMF credit/Exports (+/-)	**-1.92**	**-2.11**	**-2.09**	**-2.34**	**-2.70**	**-3.17**	**-3.02**	**6.70**	**11.20**
	(-2.56) [-2.77]	(-2.80) [-3.04]	(-2.76) [-2.97]	(-4.20)	(-4.77)	(-4.99)	(-3.25)	(4.29)	(11.7)
Vol export growth (+/-)	**2.23**	**1.97**	1.85	**3.31**	**2.93**	**2.50**	-1.55	**-4.07**	**-4.88**
	(2.60) [2.16]	(2.28) [1.90]	(2.12) [1.75]	(6.72)	(5.61)	(4.28)	(-1.84)	(-3.26)	(-3.5)
Trade balance/GDP (-)	1.23	1.05	1.05	**3.80**	**3.68**	**4.40**	**9.46**	**5.26**	-0.91
	(1.62) [1.82]	(1.37) [1.53]	(1.34) [1.49]	(6.07)	(5.74)	(6.2)	(7.67)	(3.69)	(-0.5)
Credit private /GDP (+/-)	**-3.24**	**-3.49**	**-3.55**	**-3.66**	**-3.81**	**-3.19**	**-15.61**	**-12.71**	**-16.53**
	(-5.05) [-5.46]	(-5.31) [-5.68]	(-5.32) [-5.62]	(-7.99)	(-8.19)	(-6.20)	(-12.61)	(-9.53)	(-14.81)
GDP growth (-)	-1.55	-1.24	-1.64	0.12	0.48	1.09	**-8.91**	**9.17**	**9.42**
	(-1.02) [-0.96]	(-0.81) [-0.77]	(-1.07) [-1.01]	(0.10)	(0.40)	(0.85)	(-5.16)	(3.74)	(3.87)
GNP per capita (-)	**0.49**	**0.54**	**0.56**	**0.18**	**0.26**	**-0.36**	**1.31**	**-2.30**	**0.95**
	(4.82) [4.84]	(5.18) [5.11]	(5.34) [5.29]	(2.13)	(3.00)	(-3.47)	(8.12)	(-10.27)	(3.81)
Vol GNP pc growth (+/-)	4.55	3.51	3.19	**4.07**	3.22	1.61	6.71	-7.16	**-15.09**
	(1.77) [1.78]	(1.37) [1.39]	(1.23) [1.22]	(2.16)	(1.72)	(0.79)	(1.58)	(-1.65)	(-2.52)
Real exchange rate[c] (+/-)	0.03	0.01	0.04	0.15	0.13	0.12	**0.81**	**-2.03**	**3.64**
	(0.16) [0.16]	(0.06) [0.07]	(0.22) [0.23]	(1.41)	(1.12)	(0.95)	(7.25)	(-5.86)	(11.74)
Trade/GDP (-)	**-4.70**	**-4.86**	**-5.02**	**-6.03**	**-5.90**	**-5.74**	**-16.12**	**-7.58**	**-17.79**
	(-7.96) [-7.52]	(-8.11) [-7.66]	(8.20) [-7.69]	(-13.22)	(-12.83)	(-11.7)	(-14.59)	(-7.74)	(-13.84)

The expected sign according to economic theory is in parenthesis after the variable name. The t-ratios based on the ML estimator of the asymptotic covariance matrix are in parenthesis. For the baseline PLOGIT and the FTE that do not allow for country-specific heterogeneity the t-ratios based on the Huber-White sandwich estimator are in brackets. Bold indicates significant at the 5% level.

TABLE 3.2

Logit EWS Models for Sovereign Default, 1984-2000 (h.a.c. Standard Errors)

	latent effects									AR(p)
		time		country	country, time					
Variables	PLOGIT(ng)	PLOGIT	FTE	RE(ng)	RE	RC^{γ}	RC^{γ}-AR	RC^{β}	RC^{β}-AR	$\sum_{j=1}^{p} \rho_j$
External debt/GDP (+)	**7.86**	**8.18**	**8.39**	**9.43**	**9.95**	**9.83**	**15.99**	15.25	16.20	0.535
	(9.44)	(9.75)	(11.33)	(6.98)	(7.46)	(5.61)	(2.59)	(0.77)	(0.81)	
Offic debt/ Tot debt (+)	**8.41**	**8.82**	**8.50**	**6.22**	**7.04**	6.91	8.98	11.52	18.65	0.643
	(3.47)	(3.48)	(4.09)	(2.04)	(2.26)	(1.77)	(1.43)	(0.52)	(1.37)	
ST debt/ Tot debt (+)	**5.44**	**5.41**	**5.21**	-1.31	-1.02	-1.75	2.69	1.52	8.31	0.595
	(2.54)	(2.49)	(2.83)	(-0.41)	(-0.33)	(-0.42)	(0.45)	(0.05)	(0.74)	
Debt serv/Exports (+)	**-3.42**	**-3.80**	**-4.26**	-2.69	-3.01	-2.47	-2.07	1.29	-5.72	0.454
	(-2.63)	(-2.80)	(-3.58)	(-1.71)	(-1.89)	(-1.27)	(-0.62)	(0.14)	(-0.70)	
IMF credit/Exports (+/-)	-1.92	**-2.11**	**-2.09**	-2.34	**-2.70**	**-3.17**	-3.02	**6.70**	**11.20**	0.732
	(-1.94)	(-2.15)	(-2.43)	(-1.87)	(-2.19)	(-2.13)	(-0.99)	(2.42)	(6.54)	
Vol export growth (+/-)	2.23	1.97	1.85	3.31	2.93	2.50	-1.55	-4.07	-4.88	0.620
	(1.61)	(1.41)	(1.50)	(1.61)	(1.36)	(0.97)	(-0.28)	(-0.30)	(-0.41)	
Trade balance/GDP (-)	1.23	1.05	1.05	**3.80**	**3.68**	**4.40**	9.46	5.26	-0.91	0.324
	(1.33)	(1.11)	(1.27)	(2.21)	(2.19)	(2.00)	(1.30)	(0.64)	(-0.05)	
Credit private /GDP (+/-)	**-3.24**	**-3.49**	**-3.55**	**-3.66**	**-3.81**	**-3.19**	**-15.61**	**-12.71**	**-16.53**	0.757
	(-3.63)	(-3.76)	(-4.56)	(-2.84)	(-2.88)	(-1.96)	(-2.06)	(-2.61)	(-3.16)	
GDP growth (-)	-1.55	-1.24	-1.64	0.12	0.48	1.09	-8.91	9.17	9.42	-0.122
	(-0.84)	(-0.67)	(-0.93)	(0.05)	(0.20)	(0.38)	(-1.53)	(0.57)	(0.59)	
GNP per capita (-)	**0.49**	**0.54**	**0.56**	0.18	0.26	-0.36	**1.31**	-2.30	0.95	0.853
	(3.19)	(3.41)	(4.37)	(0.97)	(1.36)	(-1.35)	(2.42)	(-0.73)	(1.19)	
Vol GNP pc growth (+/-)	4.55	3.51	3.19	4.07	3.22	1.61	6.71	-7.16	-15.09	0.630
	(1.41)	(0.89)	(0.96)	(0.78)	(0.64)	(0.25)	(0.72)	(-0.18)	(-0.87)	
Real exchange rate[c] (+/-)	0.03	0.01	0.04	0.15	0.13	0.12	0.81	-2.02	3.64	0.510
	(0.14)	(0.05)	(0.20)	(0.56)	(0.50)	(0.40)	(1.28)	(-0.58)	(1.06)	
Trade/GDP (-)	**-4.70**	**-4.86**	**-5.02**	**-6.03**	**-5.90**	**-5.74**	**-16.12**	-7.58	-17.79	0.657
	(-5.25)	(-5.45)	(6.40)	(-3.86)	(-3.96)	(-2.80)	(-2.19)	(-0.60)	(-0.94)	

The expected sign according to economic theory is in parenthesis after the variable name. The t-ratios based on the h.a.c. estimator of the asymptotic covariance matrix are in parenthesis. Bold indicates significant at the 5% level. $\sum_{j=1}^{p} \rho_j$ denotes the sum of the slope coefficients in a AR(5) model with controls for time-invariant country effects using dummies.

CHAPTER 4: Towards the Optimal Design of an Early Warning System for Debt Crises

4.1 Introduction

The financial turmoil in emerging and developing markets during the last decade has stressed the need for accurate country risk assessment. A number of studies have focused on the so-called twin crises, namely, banking and currency crises (Frankel and Rose, 1996; Berg and Pattillo, 1999; Kaminsky and Reinhart, 1999; Kumar et al., 2003).[1] As more countries are moving toward flexible exchange rates the latter are becoming less frequent events. But sovereign debt crises remain a matter of concern for international financial markets and economic policymakers.

The sovereign default literature is very prolific. Most studies have focused on identifying the main determinants of default among fundamentals of the domestic economy and indicators of the international business-cycle and market sentiment. For this purpose, different classification techniques have been used. However, scant attention has been paid to forecasting issues and to the optimal design of an Early Warning System (EWS) tailored to the decision-maker's preferences.

Various classification techniques have been attempted mostly based on macrovariables. Several studies have applied *linear discriminant analysis* which assumes multivariate normal regressors with equal covariance matrices in the default and non-default states (Frank and Cline, 1977; Taffler and Abassi, 1984). These assumptions have been shown to be rather strong. The most recent research is based on nonlinear *panel logit/probit models* (Peter, 2002). Non-parametric classification techniques such as *clustering* and *recursive tree* analysis, albeit popular in other areas, have received little attention in this literature. One exception is Manasse et al. (2003) who apply both a logit and a recursive tree

[1] See Abiad (2003) for a comprehensive survey of the EWS literature.

analysis. Moreover, most of the models thus developed are based on arbitrary choices for the *cut-off* probability and *warning horizon* or crisis window. These ad hoc choices may not necessarily be optimal for the problem at hand. For instance, a low cut-off rate and a long warning horizon may be better choices for a highly risk averse (towards default) decision maker since they lead to more default signals and vice versa.

The credit ratings provided by leading agencies and bankers have also been found to contain predictive power regarding sovereign debt crises and to Granger-cause the spreads of sovereign bonds (Reinhart, 2002; Rojas-Suárez, 2001; Larrain et al., 1997). Moreover, the New Basel Accord allows banks to use internal ratings for calculating capital requirements. The Institutional Investor ratings can be regarded as consensus internal ratings from major international banks. The upshot is that it is unclear which method and information set one should adopt to develop an EWS of sovereign default. This indirectly stresses the potential importance of forecast combining, an issue that has received scant attention in this literature.

The contribution of this chapter is twofold. First, it provides a framework for the optimal design of an EWS for sovereign default. For this purpose, it implements three forecasting tools: *i*) logit based on macrovariables, *ii*) K-means clustering of macrovariables and *iii*) logit based on Institutional Investors' credit ratings (LOGIT-R). Thesecond classifier has not been utilized in the present context as yet. The sample pertains to 75 emerging and developing economies over 1983-2000. We show how the loss function and degree of risk aversion of the decision-maker can be accounted for in order to optimally choose key elements of an EWS such as the logit cut-off rate and the number of clusters. The calibration of the classifiers is conducted in-sample recursively over a 12-year rolling window. The assessment of their forecast ability over a 5-year holdout period focuses on the anticipation of default *entry* events.

Second, the chapter delves into several forecast combining issues. It assesses the relative strengths of the classifiers for different decision-makers and the stability of the ranking over the holdout years. In order to adequately gauge the gains from forecast combining, it is shown that the choice of weighting scheme should also be tailored to the loss function and degree of risk aversion.

The remainder of the chapter is organised as follows. Section 4.2 outlines the background literature. Section 4.3 describes the methodology and Section 4.4 introduces the data. Sections 4.5 then illustrates empirical issues regarding the optimal calibration of classifiers while our forecast combining analysis is presented in Section 4.6 before concluding.

4.2 Elements in the Design of an Optimal EWS

The goal of an EWS is to issue signals of pending debt repayment difficulties. Hence, the variable of interest takes a value of one at year t if a default occurs any time within a $[t, t+h]$ window

$$y_{it} = \begin{cases} 1 & \text{if } d_{it+k} = 1 \text{ at any } k = 0, 1, ..., h-1 \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

and the classification problem at hand is formalized as $y_{it} = f(\mathbf{x}_{i,t-1})$ where $\mathbf{x}_{i,t-1}$ represents the available predictors at $t-1$. The forward-looking variable y_{it} is called the Early Warning Default (EWD) indicator.

4.2.1 Warning Horizon for a Crisis Signal

It is reasonable to expect that signs of economic deterioration will emerge some years before a debt crisis occurs. The warning horizon is the time interval within which the EWS should anticipate the occurrence of a crisis. If the warning horizon chosen is, say, $h = 3$ years then the forecast $\hat{y}_{i,t+1} = 1$ indicates that a debt crisis will occur sometime during $[t+1, t+3]$. Choosing h requires a trade-off. The longer h is the less missed defaults but the more false alarms and vice versa.

Currency crises studies typically set $h = 24$ months (Berg and Pattillo 1999; Kumar et al., 2003). Bussière and Fratzscher (2002) find empirically the optimal currency-crisis warning horizon according to a specific loss function. Burkart and Coudert (2002) carry out a sensitivity analysis of the Type I and II error rates associated to currency-crisis warning horizons from 1 to 4 quarters and find worse performance at shorter horizons (higher Type I error rate). However, the choice of warning horizon has received scant attention in the sovereign default literature. Most studies arbitrarily set $h = 1$ year. Sommerville and Taffler (1995) argue that this may be too short for bankers' purposes. Peter (2002) and Oka (2003) arbitrarily set $h = 3$ years in modeling default on external creditors and arrears to the IMF, respectively.

4.2.2 Cut-off Probability Rate

In order to assess the adequacy of an EWS, the probability forecasts are usually transformed into event forecasts and compared with the EWS indicator y_{it}. For this purpose, the decision-maker has to adopt a cut-off or threshold probability λ. Mascarenhas and Sand (1989) show that the overall error rate of a sovereign default predictor based on discriminant analysis varies with λ.

In the financial crises literature, λ is most often arbitrarily set at 0.5, 0.25 or 0.10 (Kumar et al., 2003; Peter, 2002; Berg and Pattillo, 1999; Frankel and Rose, 1996), fixed at the in-sample frequency of crises (Demirguc-Kunt and Detragiache, 1999; Detragiache and Spilimbergo, 2001; Manasse et al., 2003) or at the level that balances the Type I and II errors (Burkart and Coudert, 2002). Only three sovereign default studies account for the loss function in choosing this parameter (Taffler and Abassi, 1984; Sommerville and Taffler, 1995; Oka, 2003).

4.2.3 Decision-maker's Loss Function

The loss function facilitates the expected cost of mispredicting. On this basis, the forecaster can optimally calibrate several parameters of an EWS such as the cut-off rate and the warning horizon. The literature typically treats the Type II error (false alarms) as less worrisome than the Type I error (missed crises), mainly for two reasons. First, the costs of missing investment opportunities or those of taking pre-emptive policy measures in the case of a false warning are often less severe than the losses, reflected in the lender's balance sheet, or the welfare cost of an unanticipated default. Second, false alarms are not always 'errors' as such in that they may not stem from predictive failure of the model but simply reflect that, although economic vulnerabilities might have been severe, appropriate policy actions were taken and a debt crisis was avoided.

Suppose that an EWS is developed using the warning horizon h and the cut-off λ. On the basis of its forecasts, different error measures can be computed. Let $E_0(\lambda, h)$ and $E_1(\lambda, h)$ denote the number of false warnings ($\hat{y}_{it} = 1, y_{it} = 0$) and missed defaults ($\hat{y}_{it} = 0, y_{it} = 1$), respectively. Let $C_0(h)$ and $C_1(h)$ denote the total number of tranquil ($y_{it} = 0$) and debt crisis ($y_{it} = 1$) cases, respectively. The available sample has $C = C_0 + C_1 = NT$ cases where T denotes time periods and N denotes countries. The Type I error probability (P_I hereafter) is estimated by the percentage of missed defaults, $E_1(\lambda, h)/C_1(h)$. The Type II error probability (P_{II}) gives the likelihood of a false alarm and it can be estimated as $E_0(\lambda, h)/C_0(h)$. Finally, let θ denote the degree of risk aversion (toward missing a crisis) of the decision-maker. Below we outline three loss functions that have been widely used in the broad financial crisis literature.

Kaminsky et al. (1998) introduce the *noise-to-signal* loss (NS) for currency crisis forecasts

$$NS(\lambda, h) = \frac{P_{II}(\lambda, h)}{1 - P_I(\lambda, h)}, \quad NS \in [0, 1] \tag{4.2}$$

defined as the ratio of the probability of a false alarm over the probability of a correct crisis warning. Other currency crises studies adopt this loss function (Berg and Pattillo, 1999; Burkart and Coudert, 2002). But it has not yet been used in the sovereign default literature. It can be estimated by $\widehat{NS}(\lambda, h) = c(h) \times \frac{E_0(\lambda,h)}{C_1(h)-E_1(\lambda,h)}$ where $c(h) \equiv C_1(h)/C_0(h)$. Hence, optimizing λ according to (4.2) amounts to minimizing the ratio of false alarms to correct alarms.

Another typical loss function, that we call *investor's loss* (IL), is defined as

$$IL(\theta, \lambda, h) = \theta P_I(\lambda, h) + (1-\theta)P_{II}(\lambda, h), \quad IL \in [0, 1] \tag{4.3}$$

it can be estimated by $\widehat{IL}(\theta, \lambda, h) = \theta\frac{E_1(\lambda,h)}{C_1(h)} + (1-\theta)\frac{E_0(\lambda,h)}{C_0(h)}$. The cost attached to a missed default relative to that of a false alarm is captured by the risk-aversion parameter θ, e.g. $\theta = 0.8$ reflects that the cost ratio for the decision-maker is 4 to 1. Equation (4.3) represents a family of loss functions parameterized by θ. Oka (2003) and Burkart and Coudert (2002) adopt it for $\theta \geq 0.5$.

Other studies have employed what we call the *policymaker's loss* (PL) function defined as

$$PL(\theta, \lambda, h) = \theta P_I(\lambda, h) + (1-\theta)P_W, \quad PL \in [0, 1] \tag{4.4}$$

which is a weighted sum of the probability of missing a default and the probability of issuing an early warning. One implicit assumption is that the latter triggers some policy action or structural reform (e.g. a reduction in public-sector pay and employment) that maybe costly for policymakers in terms, for instance, of social unrest and not being reelected and for the country itself in terms of reduction in output, economic instability and so forth. Thus an optimal EWS for policymakers should not trigger too many warning signals.[2] In contrast, the loss function (4.3) presumes that correct alarms have negligible costs and so it is thought to be more representative of investors.[3]

[2]The cost of a false alarm is simply the cost of the preventive policy action $(1-\theta)$. Strictly speaking, the cost of a correct alarm is the cost of the preventive action minus the benefit from preventing the default $(1-\theta)-\zeta$.

[3]A correct default warning entails transaction costs for investors. Nevertheless, labelling (4.3) as IL

Note that $P_W \equiv \Pr(\hat{y}=1) = \Pr(\hat{y}=1 \cap y=0) + \Pr(\hat{y}=1 \cap y=1)$ so that

$$P_W = \Pr(\hat{y}=1|y=0)\Pr(y=0) + \Pr(\hat{y}=1|y=1)\Pr(y=1)$$

which can be estimated by $P_W = \frac{E_0}{C_0}\frac{C_0}{C} + (1-\frac{E_1}{C1})\frac{C_1}{C}$. Thus we have

$$\widehat{PL}(\theta,\lambda,h) = (1-\theta)\hat{p}\left\{[\frac{\theta}{(1-\theta)\hat{p}} - 1]\hat{P}_I(\lambda,h) + \frac{(1-\hat{p})}{\hat{p}}\hat{P}_{II}(\lambda,h) + 1\right\} \tag{4.5}$$

which reveals that PL is also a linear function of P_I and P_{II} but, in contrast with IL, the weights reflect not only the risk aversion θ but also the prior probability of default or in-sample default frequency $\hat{p} = \frac{C_1}{C}$. The lower $\hat{p}$ the heavier the penalty for the Type II error ceteris paribus. The PL metric is adopted by Bussière and Fratzscher (2002) and Demirguc-Kunt and Detragiache (1999). The former show how the optimal (λ, h) depends on θ whereas the latter focus on λ in the contexts of currency and banking crises, respectively.

4.2.4 Forecast Combining Schemes

Bates and Granger's (1969) seminal paper sets out the concept of forecast combination. It urges that, when alternative forecasts are available, it may pay to combine this information rather than to opt for one of the alternatives. Combination permits the blend of forecasts from a range of sources. Combination has been shown to be effective not only when the forecasts are obtained from widely heterogenous methods but more generally also (Montgomery et al., 1998; Winkler and Makridakis, 1983; Clemen et al. 1995). Instability in the ranking of forecasts provides another rationale for combination (Stock and Watson, 2001; Aiolfi and Timmermann, 2003).

Bates and Granger (1969)'s framework advocates that the gains from *quantitative* forecast combining are akin to the diversification gains in a portfolio of assets and hence, the optimal weights depend on the covariance matrix of individual forecasts. This approach

and (4.4) as PL is mainly for ease of exposition. This terminology does not imply loss of generality in the ensuing analysis.

amounts to minimising the combined forecast error variance. Granger and Ramanathan (1984) show that the latter boils down to an OLS regression of the actual series on the individual forecasts with the constraints of a zero intercept and slopes adding to one. They propose unconstrained OLS as a means of allowing for the possibility of biased forecasts. Weighted least squares has been suggested for large samples to allow for the possibility of time-varying weights and structural change (Diebold and Pauly, 1987). A Bayesian approach to forecast combining is introduced by Bunn (1975). In this spirit, Clemen and Winkler (1986) advocate shrinking the OLS weights toward some prior mean. Gupta and Wilton's (1988) Bayesian odds-matrix approach has proven quite successful for small samples. Several studies discussed in Clemen (1989) show that the simple equal-weighting approach does at least as well as the above parametric methods based on past performance. Non-parametric combination schemes — weights based on rankings of past performance — have been shown to outperform the equal-weights combination out of sample (Aiolfi and Timmermann, 2003).

The combination of forecasts for *discrete* variables raises distinct issues and this literature is comparatively smaller. Bayesian methods (Gupta and Wilton, 1988; Clemen et al., 1995) have been suggested to combine probability forecasts. In Gupta-Wilton's odds-matrix method, forecast performance is measured by the distance between the forecasted probability and the realized event. For event forecasts, Feather and Kaylen (1989) suggest an approach based on grouping them into mutually exclusive classes (method A forecasts success and method B failure, both forecast success and so forth) and conditional upon the class, the success and failure probabilities are assumed to follow a joint Dirichlet distribution. This approach has been shown to be nested in Kamstra and Kennedy's (1998) [KK] logit method. The latter is simpler to apply and it can combine probability forecasts, event forecasts and a mix. Moreover, it can be extended to polychotomous and ordered classification problems on the basis of multinomial- or ordered-logit regressions, respectively. Monte Carlo simulations in Kamstra and Kennedy (1998) and an empirical application in Kamstra et al. (2001) suggest that KK-logit can beat the simple average for large samples.

The extensive and continued interest in forecast combination is in large part explained by the wealth of evidence from empirical studies on its merits (see Newbold and Harvey, 2004). Surprisingly, this concept has received scant attention in the sovereign default literature. Sommerville and Taffler (1995) compare judgmental forecasts (bankers' ratings) and parametric forecasts (logit and linear discriminant analysis) based on macrodata but do not assess the merits of combined forecasts. Mascarenhas and Sand (1989) investigate the accuracy of discriminant analysis based on three information sets — credit ratings, macrovariables or both. They find that combining credit rating and macroeconomic forecasts using Gupta and Wilton's (1988) Bayesian odds-matrix method outperforms the individual forecasts. But they do not explore alternative weighting schemes in order to find the 'optimal' combination for the problem at hand. Manasse et al. (2003) compare the forecasts from a logit regression and a non-parametric recursive tree based on macrovariables and find that the latter yields less missed defaults but more false alarms. Using two distinct event weightings (dummies in the logit that incorporate information from the tree nodes and the unanimity principle) it is shown that forecast combining improves accuracy.

4.3 Methodology

4.3.1 Competing Classifiers

Three forecast methods are considered. First, a pooled logit model for macrovariables (LOGIT-M)

$$\log\left[\frac{p_{it}}{1-p_{it}}\right] = \alpha + \boldsymbol{\beta}'\mathbf{x}_{i,t-1},\ i = 1,...,N, t = 1,...,T \tag{4.6}$$

that implies the nonlinear relation, $p_{it} = \frac{e^{\alpha+\beta' \mathbf{x}_{i,t-1}}}{1+e^{\alpha+\beta' \mathbf{x}_{i,t-1}}}$, where $p_{it} \equiv \Pr(y_{it} = 1)$ and $\mathbf{x}_{i,t-1}$ is an $s \times 1$ vector. The coefficient, $\beta_j, j = 1, ..., s$ estimated by maximum likelihood represents the marginal effect of the macrovariable $x_{it,j}$ on the log-odds ratio $\log\left[\frac{p_{it}}{1-p_{it}}\right]$ ceteris paribus.[4]

Second, the analysis relies also on sovereign credit ratings (z_{it} hereafter) that reflect consensus bankers' judgment. Several studies have found these internal ratings to be correlated with default signals such as GDP per capita, inflation, external debt, economic development and the actual default history (Lee, 1993; Cantor and Packer, 1996). Futhermore, these credit ratings incorporate important qualitative information on default risk such as the effects of social, political and cultural conditions. A univariate logit transformation $\log\left[\frac{p_{it}}{1-p_{it}}\right] = \alpha+\beta z_{i,t-1}$ is used to generate default forecasts based on the bankers' ratings. We refer to the latter as LOGIT-R forecasts.[5]

As discussed in Section 4.2.1, a warning horizon h is embedded in the definition of y_{it}. In a logit framework, a cut-off probability is required to transform the probability estimates into EWS signals, i.e. $\hat{y}_{it} = 1$ if $\hat{p}_{it} > \lambda$ and $\hat{y}_{it} = 0$ if $\hat{p}_{it} \leq \lambda$. Hence, optimal logit calibration implies finding the (λ, h) combination that is 'best' according to the decision-maker's preferences, namely, her loss function and degree of risk-aversion θ. The following two-step optimization approach is proposed:

1) For each (θ, h) pair, compute the loss associated with the λ candidates so that $\lambda^*_{\theta h} \equiv \min_{\lambda} L(\theta, \lambda, h)$ gives the optimal cut-off rate. This facilitates a set of optimal cut-off rates denoted $\{\lambda^*_{\theta h}\}$.

2) For each θ, calculate the loss associated with $h = \{1, 2, ..., h_{\max}\}$ so that

$$h^*_\theta = \min_h [L(\theta, \lambda^*_{\theta h}, h)] \tag{4.7}$$

[4]The forecasts from the pooled logit model are shown to outperform those from more sophisticated specifications such as random coefficients logit under several loss functions (see Chapter 2).

[5]The finite-sample properties of different estimation approaches to generate rating migration probabilities from the external ratings provided by Moody's are explored in Chapter 5 by Monte Carlo simulation.

is the optimal warning horizon. This facilitates an optimal horizon and cut-off pair $(h^*_\theta, \lambda^*_{\theta h})$ for each θ. The latter is denoted (h^*, λ^*) for simplicity.

The third classification technique we employ is K-means clustering.[6] The inputs or cases are the NT observation vectors, $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, ..., x_{it,s})$, where s is the number of macrovariables. These are allocated in clusters so as to maximise within-cluster similarity and between-cluster discrepancy. The outputs are K *clusters* labelled as either default ($\hat{y} = 1$) or non-default ($\hat{y} = 0$) according to an *assignment rule.* An unseen or out-of-sample case $\mathbf{x}_{it}$ is classified as default/non-default depending on the cluster whose centroid is closer. (See Appendix 4.1). The choice of K does not follow from the algorithm and is often made subjectively. Hence, optimal clustering calibration requires finding the 'best' assignment rule and K according to the decision-maker's preferences.[7] We propose below an approach which is discussed without loss of generality for the IL function.

For a given K, the assignment rule can be optimized as follows.[8] Let $n_c(1)$ be the number of default cases (vectors $\mathbf{x}_{i,t-1}$ such that $y_{it} = 1$) in cluster c. Likewise for $n_c(0)$. Let C_1 (and C_0) denote the total number of default (non-default) cases. The loss implied by labelling cluster c as non-default is $L_{0,c}(\theta) = \theta \times P_I$ where $\hat{P}_I = \frac{n_c(1)}{C_1}$ is the estimated probability that a default case falls in cluster c. Likewise, $\hat{L}_{1,c}(\theta) = (1-\theta) \times \hat{P}_{II} = (1-\theta)\frac{n_c(0)}{C_0}$. The optimal rule for cluster c is

$$y^*_c = \underset{y \in \{0,1\}}{\operatorname{argmin}}\, L_{y,c}(\theta) \tag{4.8}$$

with loss $L^*_c(\theta)$. The minimal loss for the overall clustering is $L(\theta, K) = \sum_{c=1}^{K} L^*_c(\theta)$.[9]

Large-K clustering characterizes the sample rather well, but not necessarily the population and so it may produce poor out-of-sample forecasts. The optimal K can be found

[6] K-means clustering was chosen over hierarchical clustering techniques, such as nearest neighbour or average linkage, because for large datasets these are computationally and storagewise rather expensive.

[7] The proposed approach can be applied to different warning horizons to optimize h also.

[8] This is akin to finding the optimal cut-off rate λ^* in the logit framework.

[9] For the PL function, $L_0(c) = (1-\theta) \times P_W$ where $P_W \equiv \Pr(\hat{y} = 1)$ is estimated as the number of cases in the cluster over all sample cases, $\frac{n_c(1)+n_c(0)}{C}$. Likewise, we have $L_1(c) = \theta \times P_I$.

by a method introduced by Altman et al. (1985) to correct for overfitting bias in recursive partitioning. Consider $K \in \{2, ..., K_{\max}\}$ and define $L(\theta, K, \delta) = L(\theta, K) + \delta \times K$ where $L(\theta, K)$ is the minimal loss for a given K defined as above and $\delta \geq 0$ is an overfitting penalty. For each $\delta \in \{\delta_{1,}, .., \delta_n\}$ we find $\tilde{K}_\delta = \underset{K}{\operatorname{argmin}} L(\theta, K, \delta)$. This yields a set $\left\{\tilde{K}_{\delta_1}, ..., \tilde{K}_{\delta_n}\right\}$ from where K^* is found using a cross-validation. The sample $\{\mathbf{x}_{it}\}$ is randomly partitioned into V equally-sized groups. For say δ_1, we leave out one group and cluster the remaining cases using the above $\tilde{K}_{\delta_1}$ that was selected using the whole sample. The group cases left out are then assigned to the existing clusters. The procedure is iterated by leaving out a different group each time. The cross-validated loss associated to $\tilde{K}_{\delta_1}$ is the average loss over the V iterations $CV[L(\theta, \tilde{K}_{\delta_1})] = \frac{1}{V} \sum_{j=1}^{V} L(\theta, \tilde{K}_{\delta_1})_j$ where j denotes the group left out in iteration j. The optimal K^* minimizes the cross-validated loss

$$K^* = \underset{\tilde{K}_{\delta_j}}{\operatorname{argmin}} CV[L(\theta, \tilde{K}_\delta)], \ j = 1, 2, ..., n.$$

In this chapter, we set $K_{\max} = 10$, $\delta \in \{0.001, 0.002, ...0.01\}$ and $V = 5$.[10]

The main advantage of clustering over logit is its non-parametric nature, namely, it does not require the forecaster to formalize the relation between the exogenous macrovariables and the default event. But clustering has some pitfalls. First, it does not provide a continuous scoring scale such as the posterior probability of default and so the countries cannot be ranked in terms of default risk, which is important for international investors. Second, the main aspects of the default clusters (e.g. low trade/GDP) are often not clear-cut particularly when many variables are used and so one cannot identify the determinants of default, which is important for policymakers.

[10]Simulations have shown that $V = 5$ works well to calibrate the number of nodes in classification trees (Breiman et al., 1984). The change in IL or PL for successive K is of order 10^{-2}. The latter drives our choice of range for δ.

4.3.2 Combining the forecasts from LOGIT-M, LOGIT-R and K-clustering

Let $\{\hat{y}^m_{i,t+1}\}^M_{m=1}$ denote M rival forecasts formed at period t and $\hat{y}^C_{i,t+1} = \mathbf{R}(\hat{y}^1_{i,t+1}, ..., \hat{y}^M_{i,t+1})$ the combined forecast where $\mathbf{R}$ is a mapping or transformation. We consider two mappings: a) *logit* for mixed probability and event forecasts and b) *voting rules* for event forecasts.

Kamstra and Kennedy's (1998) [KK] logit regression method is simple to apply and it permits the combination of probability, event forecasts or a mix. It can be extended to polychotomous and ordered classification problems using multinomial or ordered logits, respectively. It has been shown that KK-logit can beat the equal-weights approach in large samples (Kamstra et al., 2001).

According to the KK-logit approach, we fit by OLS a regression of the EWS indicator (y_{it}) on a constant, the log-odds ratio forecasts ($\log\frac{\hat{p}_{it}}{1-\hat{p}_{it}}$) from LOGIT-M and LOGIT-R and the event forecasts ($\hat{y}_{it}$) from K-clustering. The coefficient estimates are the combining weights. To allow for time-variation, this approach is recursively applied in-sample over a 12-year rolling window. Thus we have weights $\mathbf{w}_\tau \equiv (w^1_\tau, w^2_\tau, w^3_\tau)$ for each set of out-of-sample forecasts, $\tau = 1996, ..., 2000$.

A nice property of KK-logit is that it enables forecast encompassing tests (Fair and Schiller, 1990). We conduct a LR test for $H_0 : w^m_\tau = 0$ for each $m = 1, 2, 3$ and the mth forecast for year τ is discarded if the statistic is insignificant. The out-of-sample combined forecasts, $\hat{p}^C_{i\tau}$, are transformed into event forecasts by means of a cut-off rate λ^*_τ. The latter is chosen optimally for each $\tau = 1996, ..., 2000$ as described in Section 4.3.1 according to the decision maker's preferences.

Event type forecasts can be combined using voting rules such as:

$$\hat{y}^C_{i,t+1} = \begin{cases} 1 & if \ \sum^M_{m=1} \omega^m_{t+1} \hat{y}^m_{i,t+1} \geq R \\ 0 & otherwise \end{cases} \tag{4.9}$$

where R is the voting parameter. There is a large literature on the properties of such combining schemes, mostly in the context of recursive trees and neural networks. The results depends on the choice of ω_{t+1}^m — weighting equally or according to cross-validated past performance — and the empirical evidence is inconclusive (Alpaydin, 1998; Ali and Pazzani, 1995). A key role is also played by R. The most frequent choices are $R = 1/2$ and $R = 1$ which, alongside the constraints $\sum_{m=1}^{M} \omega_{t+1}^m = 1$ and $\omega_{t+1}^m = \frac{1}{M}$, yield the Majority Rule (MR) and Unanimous Rule (UR), respectively. The combined forecast is thus the event predicted by the *majority* of the classifiers or by *all* of them, respectively. The empirical evidence on the relative performance of these schemes is conflicting. Battiti and Colla (1994) support the MR whereas Albanis and Batchelor (1999) support the UR. We consider both schemes.

4.4 The Dataset

The analysis is based on $N = 75$ emerging and developing countries over 1983-2000.[11] Since the predictors (explanatory variables) are lagged 1 year, the effective sample for y_{it} spans the 1984-2000 period: 1984-1995 is the initial 12-year rolling window and 1996-2000 is the holdout period.[12] The endogenous default indicator $\{d_{it}\}_{t=1984}^{2000}$ is constructed in the same spirit as in the previous chapters. The default events that follow from our definition correspond closely to those in Standard and Poor's (2001). See Appendix 4.2. The reduced set of transformed macroeconomic and financial ratios that resulted from an in-sample jacknife procedure is considered.[13] The thirteen variables thus retained are the

[11]See Appendix 4.2. The regions (number of countries parenthesis) are East Europe (7), Asia (12), Latin America (22), Middle East/North Africa (9), Africa (25). The countries are those in Chapter 2 excluding the ones for which no II credit ratings were available: Belize, Burundi, Cape Verde, Central African Rep., Chad, Eq. Guinea, Gambia, Guyana, Maldives, Mauritania, Niger, Rwanda, Sao Tome-Principe, Solomon Islands, St. Kitts-Nevis, St. Lucia,Vanuatu, Yemen.

[12]LIMDEP 8 and SPSS 10 are used in the subsequent empirical analysis.

[13]The jacknife is based on a logit regression over 1984-1995. A variable is dropped if, in doing so, the cross-validated loss (a conservative IL or PL with $\theta = 1$, $\lambda = 0.5$ and $h = 3$) does not increase.For details, see Chapter 2, Appendix 2.3. The variable selection could be cast as another element in the optimal design of an EWS, namely, one could select the set that is 'best' according to the decision-maker's preferences.

predictor set, $\mathbf{x}_{it} \equiv (x_{it,1}, x_{it,2}, ..., x_{it,13})'$, for the LOGIT-M and K-clustering classifiers.[14] The LOGIT-M estimates using the thirteen variables retained are outlined in Appendix 4.3.

Country credit ratings are obtained from the *Institutional Investors* database. These are available for 67 countries in 1984 rising to a maximum of 83 by 2000.[15] They seek to capture the perception of worldwide bankers regarding a country's ability and willingness to service its financial obligations. In particular, our ratings series (z_{it}) is an index based on the weighted scores assigned to the countries by the 100 largest international commercial banks. The latter closely monitor the observance of standards — whether a country has published an IMF Article IV or ROSC and met the SDDS specifications.[16] It varies in a 1-100 scale with 100 representing low default-risk countries. The bankers' ratings are updated semi-annually and our LOGIT-R classifier is based on end-of-year data. To avoid sample selection bias in comparing the classifiers — LOGIT-M, LOGIT-R and K-clustering — the country-period cases used in the analysis are those for which both $\mathbf{x}_{it}$ and z_{it} are available. This leads to 45 countries in $t = 1984$ and 69 in $t = 2000$.

We should stress that sovereign debt crisis typically last longer than one year in contrast with banking and currency crises. About 30% of all country-period cases over 1984-2000 are defaults ($d_{it} = 1$) whereas about 10% are default entries ($\Delta d_{it} = 1$). The average length of a debt crisis is around 3 years. The real challenge for an EWS of sovereign default is to predict a new default *entry* (turning point) rather than a perpetuating default. In order to develop a powerful EWS in the above sense, the loss functions will be

[14]The same information set is used for both classifiers to make the comparison more informative. There is also the rationale that the determinants of default shoult not depend on the model employed.

[15]In contrast, external credit ratings from Moody's and S&P's are unavailable for many countries in our sample.

[16]The Special Data Dissemination Standards (SDDS) was designed for countries with, or seeking access to, international capital markets. It sets macro data definitions, in particular, reserves. It also sets minimum timeliness and frequency standards for data releases. The Reports on the Observance of Standards (ROSC) are voluntary and refer to transparency, financial market regulation and corporate governance issues. (See Glennerster, 2004).

evaluated over an *entry* set defined as follows. Year t is excluded for country i if it was in default at year $t-1$, i.e. $d_{it} = 1$ is excluded if $d_{i,t-1} = 1$.[17]

4.5 Optimal Calibration of Forecasting Tools

This section discusses the in-sample calibration of the classifiers over the first 12-year window 1984-1995. Unless otherwise noted, y_{it} is based on $h = 1$ and the risk aversion parameter is set at $\theta = 0.5$. Asterisks denote optimal values.

4.5.1 Balancing the missed defaults and the false alarms

As noted earlier, the cut-off rate (λ) and warning horizon (h) parameters of an EWS are often chosen subjectively. An objective choice requires a trade-off between Type I and Type II errors. Figure 4.1, Panel A, illustrates the latter for the LOGIT-M classifier.[18] We consider cut-off rates $\lambda \in (0,1)$ and warning horizons $h = \{1, 2, ..., h_{\max}\}$ with $h_{\max} = 3$. A higher λ or a lower h yield fewer false alarms at the expense of more missed defaults. From the perspective of creditors or investors, the latter means realised losses (balance sheet), reserve holdings will need to rise and cash flows and asset values will also be adversely affected. From the perspective of policymakers, the experience of the 1990s has suggested that output contractions, rising unemployment and poverty rates are some of the repercussions of sovereign debt crises.

[17]Some studies use 'exclusion windows' whereby consecutive default years within a certain time window are excluded from the empirical analysis; the length of the window is arbitrary and studies have used about 1-3 years (Detragiache and Spilimbergo, 2001; Frankel and Rose, 1996). However, in using this reduced sample for the model specification and estimation one may discard important information. In this paper, the *entry* set is used to assess the models' out-of-sample forecast accuracy. The estimation and clustering are based on all the in-sample cases.

[18]The LOGIT-R calibration raises similar issues to that of the LOGIT-M so we focus on the latter. Clustering methods have been shown to work better when the variables are mapped to the [0,1] interval. Hence, for K-clustering the N points for each year, $\{x_{it}\}_{i=1}^{N}$, are transformed using $\tilde{x}_{it} = (x_{it} - \min\{x_{it}\})/(\max\{x_{it}\} - \min\{x_{it}\})$. The same 13 regressors are used for both the LOGIT-M and K-clustering classifiers to make the comparison more informative. Another rationale is that the determinants of default should not depend on the model employed.

Lowering λ or raising h will induce the opposite trade-off, namely, less missed crises at the expense of more false alarms. The latter means foregone profit opportunities for investors and unnecessary policy actions which may be costly, for instance, in terms of social unrest. Forecasters should use the (λ, h) combination that are 'best' according to the decision-maker's preferences.

4.5.2 Well-behaved Loss Functions

Figure 4.1, Panel B, shows the interaction between the losses and the cut-off $\lambda \in (0,1)$. The NS loss function monotonically falls as λ increases. The minimum NS is achieved at a relatively high $\lambda^* \geq 0.724$. The total number of default signals, $E_0 + (C_1 - E_1)$ at this λ^* is rather small and the model misses most of the defaults ($\hat{P}_I = E_1/C_1 = 85\%$). The intuition is that, as λ increases, both the rate of false alarms, E_0/C_0, and the rate of correct default warnings, $(C_1 - E_1)/C_1$, fall but the former does so faster because $C_0 > C_1$.[19] This suggests that the NS loss function may be unsuitable in this context because it leads to a debt-crisis EWS with a very high probability of missed defaults.

The main pitfall of NS is that it only accounts for the Type I and II error rates in relative terms. For instance, $NS = 1/9$ could stem from $\hat{P}_{II} = \frac{E_0}{C_0} = 10\%$ and $\hat{P}_I = \frac{E_1}{C_1} = 10\%$ or from $\hat{P}_{II} = 1\%$ and $\hat{P}_I = 91\%$. Oka (2003) and Mulder et al. (2002) pose a similar criticism for the NS in the context of arrears to the IMF and currency crises, respectively. Moreover, the NS loss function, in contrast to IL and PL, does not allow the forecaster to control for the decision-maker's degree of risk aversion (θ) in the design of an EWS. Hence, we shall focus on IL and PL hereafter.

4.5.3 Optimal Cut-off and Warning Horizon Combination

Which are the optimal warning horizon and cut-off for an EWS of sovereign default? To answer this question, we deploy the optimization approach suggested in Section 4.3.1

[19] Over the 1984-1995 period, the probability of default *entry* is $\hat{p} = C_1/(C_1 + C_0) = 71/423 = 17\%$.

for $h \in \{1, 2, 3\}$ and $\lambda \in [0.17, 1)$.[20] Figure 4.2 illustrates the in-sample calibration of LOGIT-M for representative risk-affine ($\theta = 0.2$), risk-neutral ($\theta = 0.5$) and risk-averse ($\theta = 0.8$) decision-makers under the IL or PL loss function. Table 4.1 sets out the results. The entries are the optimal values. For the IL function, a risk-neutral decision maker ($\theta = 0.5$) would need $\lambda^* = 0.641$ and $h^* = 3$. The best design for a highly risk-averse user ($\theta = 0.8$) corresponds to $\lambda^* = 0.205$ and $h^* = 2$ whereas for the risk-affine user ($\theta = 0.2$) the optimal choice is $\lambda^* = 0.727$ and $h^* = 3$. Regarding the PL function, for the risk-neutral user ($\theta = 0.5$) the best parameters are $\lambda^* = 0.270$ and $h^* = 1$ whereas for the risky user ($\theta = 0.2$) we have $\lambda^* = 0.724$ and $h^* = 1$. These findings illustrate that:

Result 1. *The optimal cut-off and warning horizon parameters of a logit EWS depend on the decision-maker's preferences (loss function, risk aversion).*

Figure 4.3, Panel A, shows the optimal cut-off for different risk aversion (towards default) levels. For a given horizon h, a higher θ needs a lower λ^* to achieve the best balance between the two errors. Hence, more risk-averse decision-makers would need lower cut-off rates. Thus we have:

Result 2. *For a given warning horizon, the optimal cut-off rate decreases with the decision-maker's degree of risk-aversion towards missing defaults.*

Figure 4.3(A) reveals also that for a given θ, a longer h implies a higher λ^* and vice versa. The intuition is that the longer the warning horizon h ceteris paribus, the more alarms are issued. The latter implies more Type II errors ($E_0 \uparrow$) but less Type I errors ($E_1 \downarrow$) with the former increasing at a faster rate and so a higher cut-off is needed to achieve the desired balance. Thus we have:

Result 3. *For a given risk aversion level, the optimal cut-off increases with the warning horizon.*

[20] The LOGIT-M for $h = \{2, 3\}$ and $\lambda < 0.17$ predicts 1 nearly all the time. The step size for λ is 10^{-4}.

The relation between the risk aversion level and the warning horizon is shown in Figure 4.3, Panel B. For any given $\theta \leq 0.7$, the IL function yields $h^* = 3$. For $\theta > 0.7$, it results in $h^* = \{2, 3\}$. Intuitively, the benefit for institutional investors from using a relatively long horizon stems from missing less defaults ($P_I \downarrow$) which outweighs the opportunity cost of more false alarms ($P_{II} \uparrow$).

Result 4. *The optimal warning horizon is relatively long for international investors.*

According to the PL function, for risk aversion $\theta \leq 0.7$ the logit EWS should be based on $h^* = 1$, for $0.75 \leq \theta \leq 0.85$ the best choice is $h^* = 2$ whereas for $\theta \geq 0.9$ it is $h^* = 3$. Thus we have:

Result 5. *The optimal warning horizon increases with the degree of risk-aversion for policymakers.*

In sum, the optimal horizon depends on the decision-maker's preferences. The contrast between Result 4 and 5 can be rationalized as follows. Since policymakers assign a cost to default alarms (whether correct or false), a longer horizon would only be optimal for those policymakers with high risk-aversion so that the benefit of missing less defaults outweighs the cost of too many alarms.

4.5.4 Optimal Number of Clusters

We now turn to the K-means clustering classifier. We consider $K \in \{2, 3, ..., K_{\max}\}$ with $K_{\max} = 10$. The optimal number of clusters under both the IL and PL functions with risk aversion $\theta \in \{0.3, 0.5, 0.8\}$ are set out in Table 4.1.[21] The optimization results are very similar under both loss functions. Figure 4.4(A) illustrates the calibration under the IL function with $\theta = 0.5$. For the low risk aversion level $\theta = 0.3$, the optimal

[21] We adopt $\theta = 0.3$ as *low* risk aversion level here because under $\theta = 0.2$ all clusters are labelled as non-default.

number of clusters is $K^* = 8$ whereas for $\theta = 0.5$ and $\theta = 0.8$ we have $K^* = 7$ and $K^* = 6$, respectively. Figure 4.4(B) also shows the relation between the optimal number of clusters and degree of risk aversion which, interestingly, is roughly V-shaped. A similar relation is found under the PL function. The main finding is

Result 6. *The optimal number of clusters depends on the decision-maker's degree of risk-aversion.*

The calibration of the warning horizon in K-clustering raises similar issues as in the logit. For instance, a higher θ leads to a longer optimal h in both classifiers.[22] The choice of assignment rule for the final clusters into default/non-default is akin to the choice of cut-off in the logit. For a high θ, the preferred assignment rule assigns relatively more clusters to the default state.

4.6 Optimal Forecast Combination

The above results suggest that the decision-maker's preferences should be accounted for in the design of an EWS, namely, in choosing parameters such as the warning horizon, cut-off rate (logit) and number of clusters (K-clustering). Another potential way to improve the performance of an EWS is by combining the strengths of different classifiers. Since many weighting schemes are possible, we suggest to choose among them optimally according to the decision-maker's preferences. To simplify the exposition, we take y_{it} as defined in (4.1) for $h = 1$ as the event to be forecasted and focus primarily on the IL function.[23] The results for the PL function are outlined at the end of this section.

[22] In contrast with the logit estimation, changing h (or the definition of y_{it}) does not change the clustering of the $\mathbf{x}_{i,t-1}$ cases. However, changing h will affect the optimal K and assignment rule for a given loss function.

[23] The same warning horizon has to be adopted in all classifiers because, otherwise we would be combining forecasts for different dependent variables (y_{it}). The calibration of h for the combined classifier can be carried out as in Section 5, namely, by choosing $h \in \{1, 2, ..., h_{\max}\}$ so as to minimize the overall loss of the combined forecasts.

We start by comparing the out-of-sample predictive ability of the classifiers. First, the classifiers are calibrated — the cut-off for LOGIT-M and LOGIT-R and the number of clusters and assignment rule for clustering — over the 1984-1995 window. The logit estimates and final clusters thus obtained are used to generate out-of-sample forecasts for 1996. This calibration and estimation/clustering is reconducted over 1985-1996 to generate out-of-sample forecasts for 1997 and so on. A country-mean loss is obtained for each out-of-sample year and then averaged over years.

Table 4.2 sets out the comparison across individual classifiers over the holdout sample. The entries are the Type I and II error rates and the overall loss (IL) for several risk aversion levels θ. The Type I error rate from LOGIT-R is lower than that from LOGIT-M virtually for all θ. The exceptions are $\theta = 0.2$ and marginally $\theta = 0.6$. So LOGIT-R outperforms LOGIT-M regarding missed default entries whereas the opposite holds for false alarms. This indirectly suggests that the bankers' judgments implicit in the ratings are relatively pessimistic about country creditworthiness.

The Type I error rate of K-clustering is essentially higher than that of LOGIT-R or LOGIT-M for low to moderate risk aversion levels $\theta < 0.55$. For higher θ, clustering gives few missed defaults (2-6%) albeit at the expense of many false alarms (69-88%). LOGIT-M classifies the non-defaults relatively well, that is, it dominates the other classifiers in terms of the Type II error rate.[24]

The ranking of the classifiers, in terms of the overall loss (IL), follows from their different relative strengths in terms of Type I and Type II errors. For instance, at the low level $\theta = 0.3$ the minimal loss is that of the LOGIT-M because of its relatively small Type II error rate despite having a large Type I error rate. However, as the risk aversion (Type I error penalty) increases, the LOGIT-R beats the LOGIT-M.[25] The overall loss of the

[24] The exceptions are $\theta = \{0.2, 0.6\}$ for which the credit ratings (LOGIT-R) yield the smallest Type II error rate.

[25] The only exceptions occur at $0.65 \leq \theta \leq 0.7$.

non-parametric (clustering) classifier is relatively large except for very high risk aversion levels $\theta \geq 0.85$. These considerations prompt the thought that there may be gains from combining the forecasts of the three classifiers.

We now assess the stability of the out-of-sample forecast ranking. Table 4.3 indicates year-by-year the best classifier and the associated minimal loss for several risk-aversion levels. The forecast ranking changes over time which further motivates the forecast combination. For instance, LOGIT-R stands out over 1996, 1998 and 2000 whereas LOGIT-M essentially excels over 1997. The forecast instability pattern can be explained in terms of the relative strengths (Type I and Type II errors) of the classifiers. In particular, LOGIT-R is relatively 'pessimistic' toward country creditworthiness and so it does quite well in 1996 and 1998 where a relatively large number of defaults occurred. Interestingly, relatively few default entries occurred in 1997.

To combine the forecasts we use the (parametric) KK-logit regression and the two non-parametric schemes — the majority rule (MR) and unanimous rule (UR). In contrast to the latter, KK-logit accounts for the historical (in-sample) forecast performance of the rival forecasts. Table 4.4 reports the KK-logit weights of the out-of-sample forecasts for year 1996. These combining weights are obtained via a logit regression of the indicator $\{y_{it}\}_{t=1984}^{1995}$ on the rival forecasts from LOGIT-M, LOGIT-R and K-clustering. The latter change with the optimal number of clusters and assignment rule which, in turn, depend on the decision-maker's risk aversion θ as the foregoing analysis has shown. Thus a new set of K-clustering forecasts is obtained as θ varies and so the forecast combining weights vary also. The cut-off rates (for the LOGITs) play no role in this combining exercise because $\hat{p}_{it}$ is directly used.

The properties of the combined forecasts are set out in Table 4.5. Regarding false alarms, the best results stem from the UR for nearly all risk aversion levels. This is expected given that, in order to signal a default, the UR requires all individual classifiers to predict a default. In terms of missed defaults, KK-logit excels for low risk-aversion levels $\theta \leq 0.45$ whereas MR leads the race for $\theta > 0.5$.[26] The UR scheme performs rather poorly relative to individual or rival combined forecasts in terms of missed defaults.

Regarding the overall loss (IL), the best combined forecasts stem either from KK-logit for risk-aversion $\theta \leq 0.75$ or from MR for $\theta > 0.75$. We conduct a Diebold-Mariano (1999) [DM] test to compare the best combined forecaster and the best individual forecaster.[27] For several risk aversion levels, it pays to combine the classifiers. For instance, for $\theta \in \{0.2, 0.25, 0.55, 0.65, 0.7, 0.75\}$ the minimal loss from KK-logit combining is significantly smaller than the losses from either of the individual classifiers. For $\theta \in \{0.55, 0.85\}$ the gains from MR combining are significant. Figure 4.5, Panel A, illustrates the comparison graphically. It shows that either the individual LOGIT-R forecasts or the combined forecasts using KK-logit or MR produce the best out-of-sample forecast performance.[28]

The forecasts should also be compared with naive predictions. Our uninformative, naive model predicts 1 for highly risk-averse decision-makers, $\theta > 0.5$, 0 for $\theta < 0.5$ and the most frequently observed event in-sample (here 0) for $\theta = 0.5$. Table 4.6 reports the ratio of the overall loss for each classifier (IL) relative to that of the naive predictor (IL^n). A DM test is conducted to compare the minimal loss among the classifiers, individual or combined, with that of the naive predictor. It turns out that for all risk-aversion levels the best forecasting model significantly outperforms the naive predictor. The highest gains relative to the naive come from LOGIT-R for $\theta = 0.5$ with the smallest ratio $\frac{\text{IL}}{\text{IL}^n} = 0.453$, followed by the KK-logit or MR combined forecasts for $\theta = 0.55$ with a ratio of

[26]The Type I and Type II error rates for the KK-logit forecasts essentially stabilize for $\theta \geq 0.6$.

[27]We compute $DM_t, t = 1, ..., m$, where $DM_t \overset{a}{\sim} N(0,1)$. Under independence between the test statistics, it follows that $DM = \frac{1}{m}\sum_{t=1}^{m} DM_t \overset{a}{\sim} N(0, \frac{1}{m})$.

[28]For the baseline *overall error* rate reported in many studies — errors over sample cases $(E_0+E_1)/C$ — the UR forecasts essentially beat all other forecasts despite their relatively high Type I error rate (E_1/C_1). This is because this metric does not account for E_1/C_1. Due to the small number of 1s in the sample, C is much larger than C_1 and so E_1/C and E_0/C appear small relative to E_1/C_1. Hence, like the NS ratio this metric can be misleading.

0.462. Interestingly, the gains of the best-performing model relative to the naive $(1 - \frac{\text{IL}}{\text{IL}^n})$ monotonically increase with θ up to $\theta = 0.5$ and then decrease thereafter.

Next we reconduct the above steps in the optimal EWS design (calibration, estimation/clustering, forecast combination and evaluation) on the basis of the PL function. The comparison of individual and combined forecasts over the holdout sample is set out in Figure 4.5, Panel B. The results suggest that forecast combining brings gains for a wide range of risk-aversion levels albeit not for all. More specifically, the KK-logit combined forecasts achieve the minimal loss for $\theta < 0.7$ whereas the LOGIT-M and LOGIT-R forecasts are ranked best for $0.7 \leq \theta \leq 0.8$. Clustering significantly outperforms all other individual and combined methods for $\theta > 0.8$.

4.7 Conclusion

This chapter highlights the importance of and illustrates how to optimally design an EWS for sovereign default according to the decision-maker's preferences. Debt crisis forecasts are obtained from two different methods based on the same macrovariables — a logit regression (LOGIT-M) and clustering — and a logit regression based on bank-internal ratings (LOGIT-R). The data pertains to 75 emerging and developing economies over 1983-2000.

First, the study shows how to recursively calibrate in-sample these classifiers according to the decision-maker's preferences. The latter are formalized by means of a loss function and risk aversion parameter. Second, we discuss forecast combining issues. For this purpose, we consider a regression framework that exploits information on the classifiers' past forecast ability and two non-parametric voting rules based on equal weights.

The results suggest that the decision-maker's preferences influence the optimal warning horizon, cut-off probability, assignment rule and number of clusters. These key parameters have mostly been chosen in an ad hoc manner in the literature. The optimally calibrated classifiers show different strengths in terms of missed defaults and false alarms. In particular, the LOGIT-M classifier outperforms the non-parametric (clustering) and judgmental (LOGIT-R) classifiers in terms of false alarms. On the other hand, judgmental and non-parametric classifiers dominate LOGIT-M in terms of missed defaults. Moreover, there is instability in the out-of-sample forecast ranking. Overall these findings vindicate a forecast combining exercise. The latter reveals that the best combining scheme depends on the decision-maker's preferences. In most cases, the combined forecasts significantly outperform the individual forecasts and uninformative naive forecasts.

The findings in this chapter should have strong implications in applied work on credit risk prediction. In practice, many EWS for sovereign default are based on ad hoc parameter choices. The optimal recursive in-sample calibration of the forecasting tools that underlie such EWS, including the forecast-combining weigthing scheme, is strongly recommended.

TABLE 4.1
Optimal Calibration of Classifiers over Estimation Window

Risk-aversion	LOGIT-M				K-clustering			
	IL		PL		IL		PL	
θ	(λ^*, h^*)	L^*	(λ^*, h^*)	L^*	K^*	L^*	K^*	L^*
0.2	(0.727, 3)	0.129	(0.724, 1)	0.195	8	0.193	8	0.288
0.5	(0.641, 3)	0.227	(0.270, 1)	0.301	7	0.343	7	0.369
0.8	(0.205, 2)	0.139	(0.205, 2)	0.161	6	0.189	6	0.196

Investors' loss (IL) or policymakers' loss (PL) with risk-aversion parameter θ. The optimal cut-off and warning horizon (λ^*, h^*) or number of clusters (K^*) give the minimal loss L^* over 1984-1995. K-clustering results based on the optimal assignment rule of final K clusters into 1 or 0 for $y_{it}(h{=}1)$.

TABLE 4.2

Out-of-sample Forecast Ability of Competing Classifiers

Classifier	Risk aversion parameter (θ)															
	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
A: *Type I error* (missed defaults)																
LOGIT-M	**0.844**	0.824	0.730	0.667	0.667	0.607	0.443	0.243	0.137	0.137	0.137	0.137	0.109	0.089	0.089	0.060
LOGIT-R	1.000	**0.779**	**0.721**	**0.564**	**0.357**	**0.207**	**0.187**	**0.187**	0.139	0.099	0.070	0.070	0.070	0.070	0.020	0.020
Clustering	0.971	0.901	0.901	0.687	0.786	0.603	0.493	0.423	**0.060**	**0.060**	**0.060**	**0.060**	**0.060**	**0.020**	**0.020**	**0.020**
B: *Type II error* (false alarms)																
LOGIT-M	0.031	**0.035**	**0.042**	**0.070**	**0.070**	**0.084**	**0.205**	**0.214**	0.350	**0.367**	**0.367**	**0.449**	**0.520**	**0.572**	**0.572**	**0.773**
LOGIT-R	**0.009**	0.048	0.061	0.100	0.131	0.239	0.266	0.266	**0.337**	0.494	0.564	0.564	0.582	0.669	0.788	0.833
Clustering	0.026	0.086	0.086	0.128	0.069	0.190	0.337	0.408	0.690	0.683	0.794	0.794	0.794	0.834	0.834	0.877
C: *Overall loss* (IL)																
LOGIT-M	**0.193**	0.232	**0.248**	0.279	0.309	0.319	0.324	0.230	0.222	**0.218**	**0.206**	0.215	0.191	0.161	0.137	0.096
LOGIT-R	0.207	**0.231**	0.259	**0.262**	**0.221**	**0.227**	**0.227**	**0.223**	**0.218**	0.237	0.218	**0.194**	**0.172**	0.160	**0.097**	**0.061**
Clustering	0.215	0.290	0.330	0.324	0.356	0.376	0.415	0.416	0.312	0.278	0.208	0.243	0.207	**0.142**	0.101	0.063

The reported Type I error is $\hat{P}_I = \frac{E_1}{C_1}$, the Type II error is $\hat{P}_{II} = \frac{E_0}{C_0}$, the overall loss is $\widehat{IL}(\theta) = \theta\hat{P}_I + (1-\theta)\hat{P}_{II}$. E_i and C_i (event i=0,1) are the number of prediction errors and sample cases, respectively. All metrics are evaluated over the holdout entry sample 1996-2000. The out-of-sample forecasts are generated recursively from classifiers calibrated over a 12-year rolling window. Bold denotes the best outcome.

TABLE 4.3

Stability of Forecast Ranking over Holdout Period

	Risk aversion parameter (θ)															
Year	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
1996	Cluster	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-M	LOG-R	LOG-R	LOG-R	LOG-R	LOG-M
	(0.189)	(0.138)	(0.148)	(0.158)	(0.124)	(0.272)	(0.260)	(0.249)	(0.240)	(0.210)	(0.187)	(0.172)	(0.138)	(0.103)	(0.069)	(0.041)
1997	LOG-M	LOG-R	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-R
	(0.167)	(0.203)	(0.120)	(0.129)	(0.138)	(0.159)	(0.170)	(0.153)	(0.136)	(0.119)	(0.102)	(0.154)	(0.123)	(0.093)	(0.062)	(0.034)
1998	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R
	(0.198)	(0.217)	(0.256)	(0.295)	(0.268)	(0.263)	(0.284)	(0.302)	(0.280)	(0.240)	(0.247)	(0.223)	(0.198)	(0.191)	(0.160)	(0.130)
1999	Cluster	Cluster	Cluster	LOG-R	LOG-R	LOG-R	LOG-R	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-M	LOG-R
	(0.200)	(0.250)	(0.300)	(0.244)	(0.277)	(0.140)	(0.128)	(0.048)	(0.102)	(0.112)	(0.096)	(0.080)	(0.064)	(0.073)	(0.049)	(0.043)
2000	LOG-M	LOG-M	LOG-M	Cluster	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	LOG-R	Cluster	LOG-R
	(0.160)	(0.211)	(0.245)	(0.178)	(0.135)	(0.160)	(0.191)	(0.186)	(0.181)	(0.192)	(0.137)	(0.114)	(0.109)	(0.098)	(0.076)	(0.040)

The Investor's Loss (IL) metric is evaluated over the holdout ***default entry*** sample 1996-2000. For each out-of-sample year we indicate the best model and the minimal $\widehat{IL}(\theta) = \theta\hat{P}_I + (1-\theta)\hat{P}_{II}$ is reported in parenthesis. The out-of-sample forecasts are generated recursively from classifiers calibrated over a 12-year rolling window. LOG-M and LOG-R denote the logit classifier based on macrovariables and credit ratings, respectively. Cluster denote the K-means clustering classifier.

TABLE 4.4
Combination Weights Based on KK-logit Regression

Combining	Risk-aversion parameter (θ)									
weights (ω_τ^m)	0.2	0.25-0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7-0.8	0.85-0.95
LOGIT-M	0.985	1.016	0.989	0.959	0.977	0.959	1.002	1.007	1.005	0.991
	(5.66)	(5.61)	(5.69)	(5.65)	(5.72)	(5.65)	(5.85)	(5.86)	(5.90)	(5.77)
LOGIT-R	0.591	0.605	0.577	0.579	0.577	0.579	0.573	0.590	0.565	0.614
	(3.39)	(3.47)	(3.24)	(3.36)	(3.33)	(3.36)	(3.17)	(3.28)	(3.08)	(3.54)
Clustering	0.390	-0.023	0.204	0.672	0.400	0.672	0.303	0.168	0.357	27.304
	(0.74)	(-0.05)	(0.60)	(2.03)	(1.26)	(2.03)	(0.566)	(0.29)	(0.61)	(0.00)
Intercept	-0.395	-0.330	-0.430	-0.817	-0.614	-0.817	-0.628	-0.500	-0.685	-27.64
	(-1.90)	(-1.53)	(-1.72)	(-2.64)	(-2.08)	(-2.64)	(-1.14)	(-0.82)	(-1.12)	(0.00)

The results are for the $\tau = 1996$ forecasts. In parenthesis, the t-ratio for the coefficient significance.

TABLE 4.5

Forecast Ability of Combined Classifiers

Weighting	Risk aversion parameter (θ)															
scheme	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
A: *Type I error* (missed defaults)																
UR	1.000	1.000	1.000	0.921	0.893	0.803	0.729	0.579	0.227	0.387	0.207	0.207	0.179	0.139	0.089	0.060
MR	1.000	0.893	0.843	0.621	0.689	0.486	**0.257**	**0.157**	**0.069**	**0.049**	**0.040**	**0.040**	**0.040**	**0.020**	**0.020**	**0.020**
KK-logit	**0.710**	**0.681**	**0.681**	**0.563**	**0.563**	**0.443**	0.343	**0.157**	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.109
B: *Type II error* (false alarms)																
UR	**0.000**	**0.004**	**0.004**	**0.022**	**0.004**	**0.017**	**0.007**	**0.095**	**0.215**	**0.262**	**0.293**	**0.345**	**0.398**	0.446	0.498	0.671
MR	0.009	0.017	0.026	0.047	0.044	0.141	0.261	0.270	0.411	0.521	0.590	0.616	0.646	0.734	0.786	0.881
KK-logit	0.044	0.048	0.061	0.075	0.070	0.101	0.145	0.270	0.367	0.380	0.393	0.402	0.419	**0.419**	**0.419**	**0.419**
C: *Overall loss* (IL)																
UR	0.200	0.253	0.303	0.337	0.360	0.371	0.399	0.361	0.222	0.343	0.233	0.241	0.222	0.185	0.130	0.091
MR	0.207	0.236	0.271	0.248	0.302	0.296	0.259	**0.208***	**0.206**	0.214	0.205	0.184	**0.161**	**0.127***	**0.097**	**0.063**
KK-logit	**0.177***	**0.207***	**0.247**	**0.245**	**0.267**	**0.255**	**0.244**	**0.208***	0.212	**0.204***	**0.194***	**0.182***	0.171	0.155	0.140	0.124

See footnote to Table 2. The forecast ability (based on the IL function) of the best combined classifier is compared with that of the best individual classifier (reported in Table 2) using a Diebold-Mariano test. ** and * denote significant at the 1% and 5% level, respectively. UR and MR are the unanimous and majority 'voting' rule, respectively. KK-logit is the regression based scheme.

TABLE 4.6

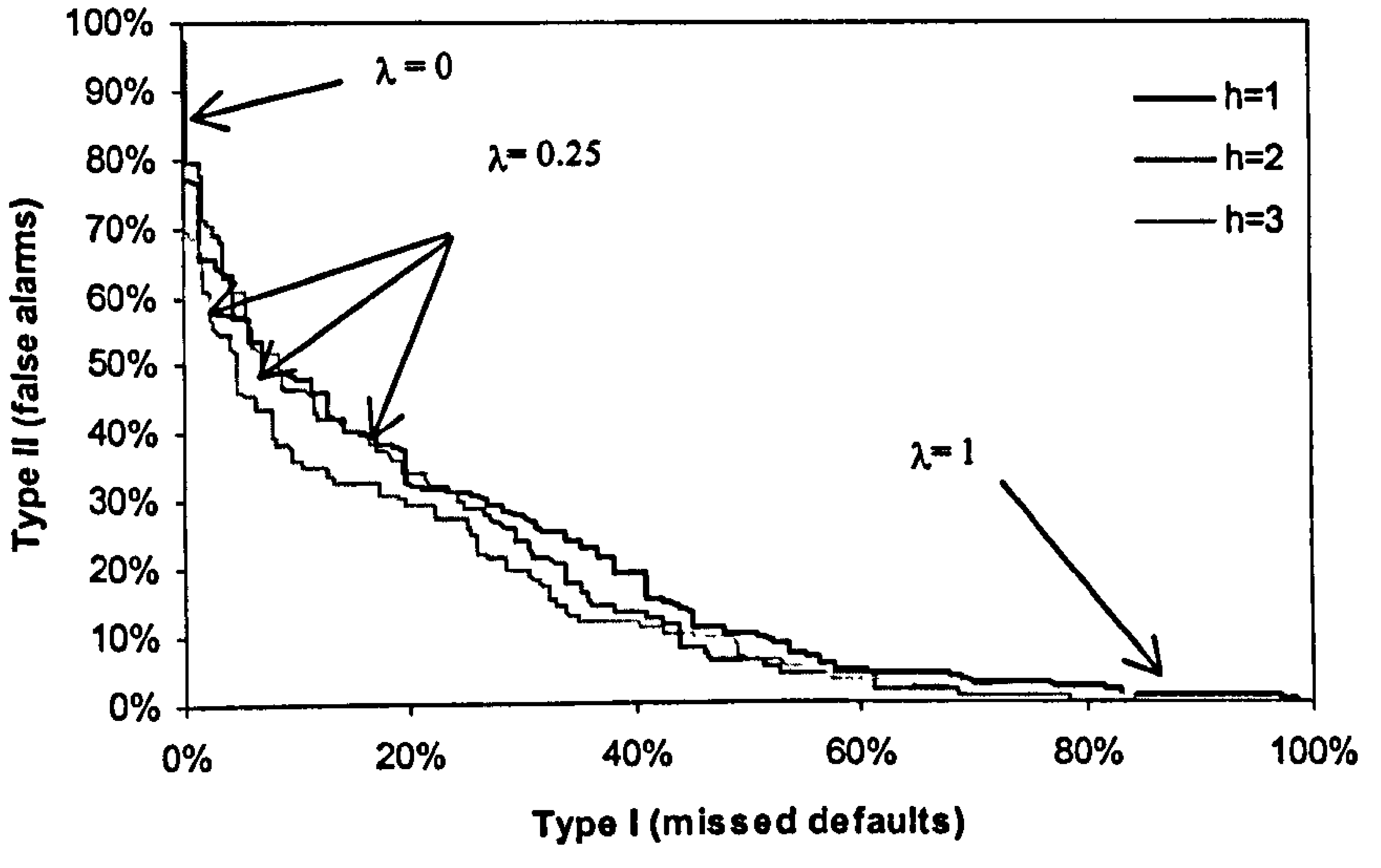
Ratio of the Classifier's Loss to the Loss of the Naive Predictor

	Risk aversion parameter (θ)															
Classifiers	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
A: *Individual*																
LOGIT-M	0.967	0.930	**0.826**	0.798	0.773	0.709	0.648	0.511	0.555	**0.622**	**0.687**	0.860	0.954	1.074	1.369	1.913
LOGIT-R	**1.036**	**0.924**	0.865	**0.750**	**0.553***	**0.500***	**0.453***	**0.495**	**0.545**	0.677	0.728	**0.774**	**0.862**	**1.065**	**0.968**	**1.213**
Clustering	1.074	1.158	1.101	0.925	0.889	0.835	0.830	0.925	0.780	0.794	0.934	0.974	1.034	0.947	1.014	1.257
B: *Combined*																
UR	1.000	1.013	1.010	0.962	0.899	0.824	0.798	0.802	0.555	0.981	0.776	0.966	1.112	1.232	1.295	1.811
MR	1.035	**0.827***	0.904	0.709	0.754	0.658	0.518	**0.462***	**0.514***	0.611	0.683	0.736	**0.806***	**0.848***	**0.966***	**1.261**
KK-logit	**0.886***	**0.827***	**0.745***	**0.701***	**0.668**	**0.566**	**0.487**	**0.462***	0.530	**0.582***	**0.646***	**0.727***	0.853	1.034	1.396	2.482

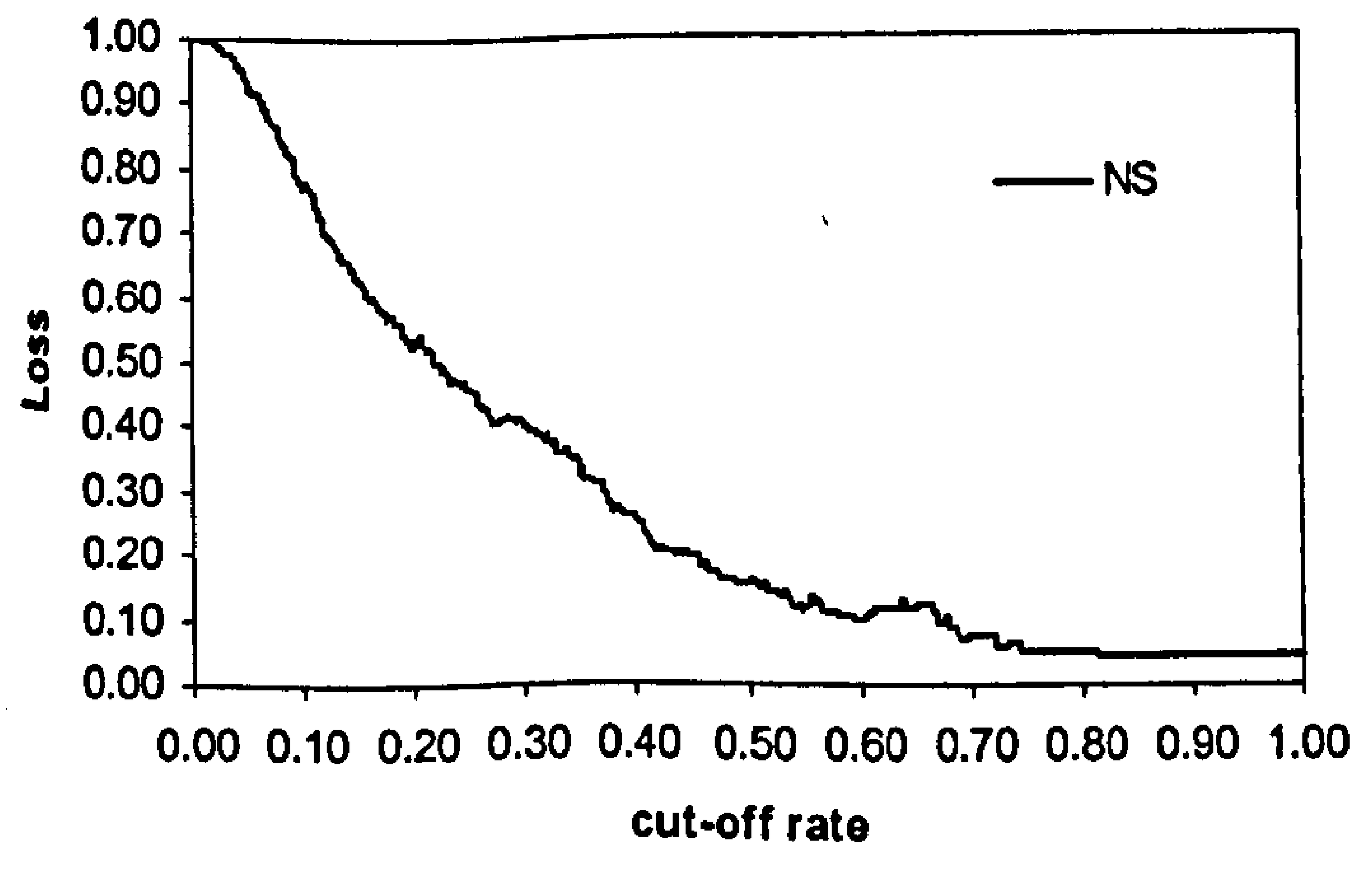
*** indicates that the overall accuracy (IL) of the best forecasts, individual or combined, is significantly better than that of the naive model at the 1% level on the basis of a Diebold Mariano test. The naive forecast is 0 for θ<0.5, 1 for θ>0.5 and the in-sample most frequent event for θ=0.5.**

Figure 4.1: The Type I and Type II errors and the Cut-off Rate

Panel A: The Trade-off Between the Probability of Type I and Type II Errors (LOGIT-M)



Panel B: The Overall Loss and the Cut-off Rate (LOGIT)



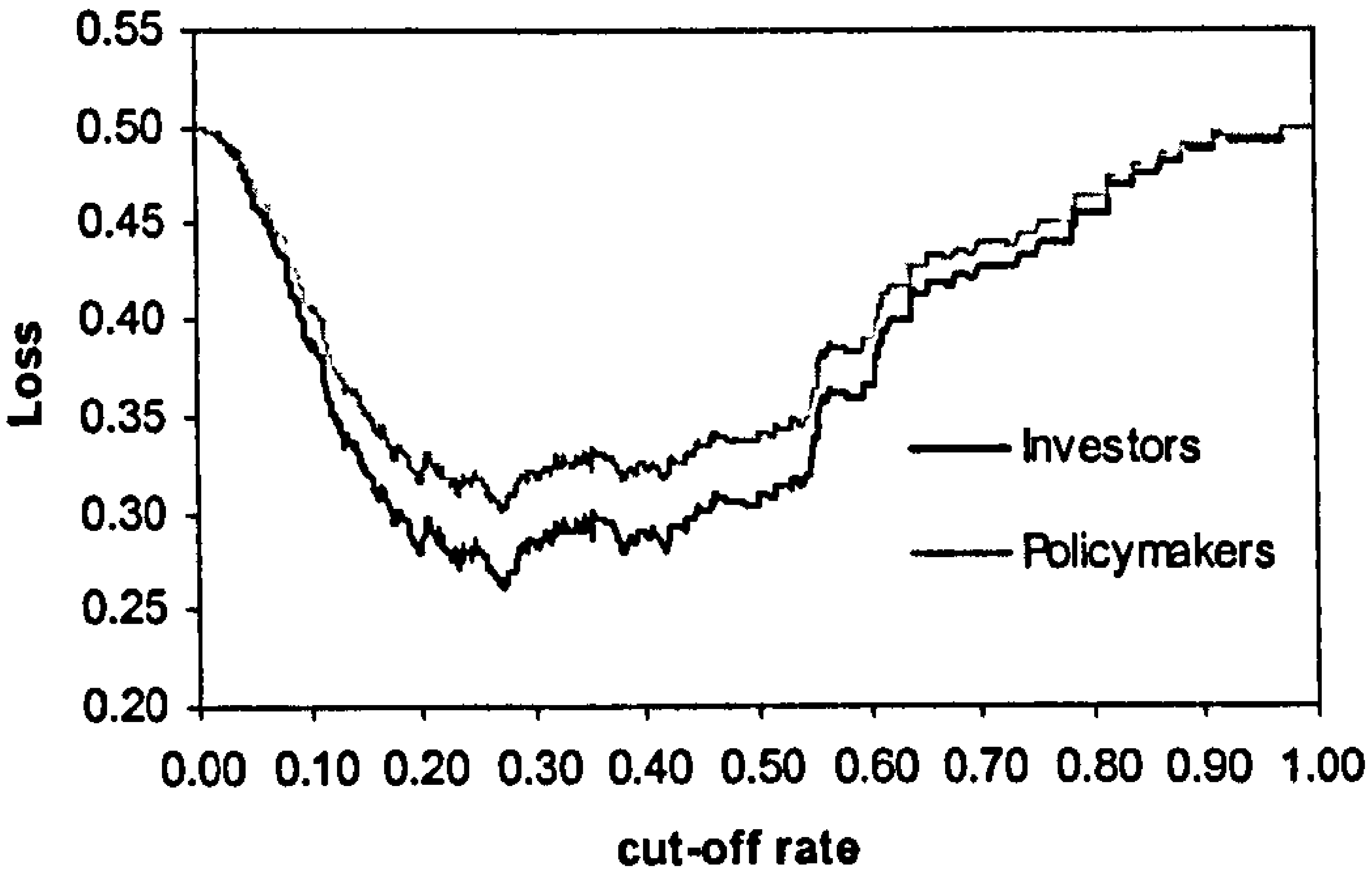
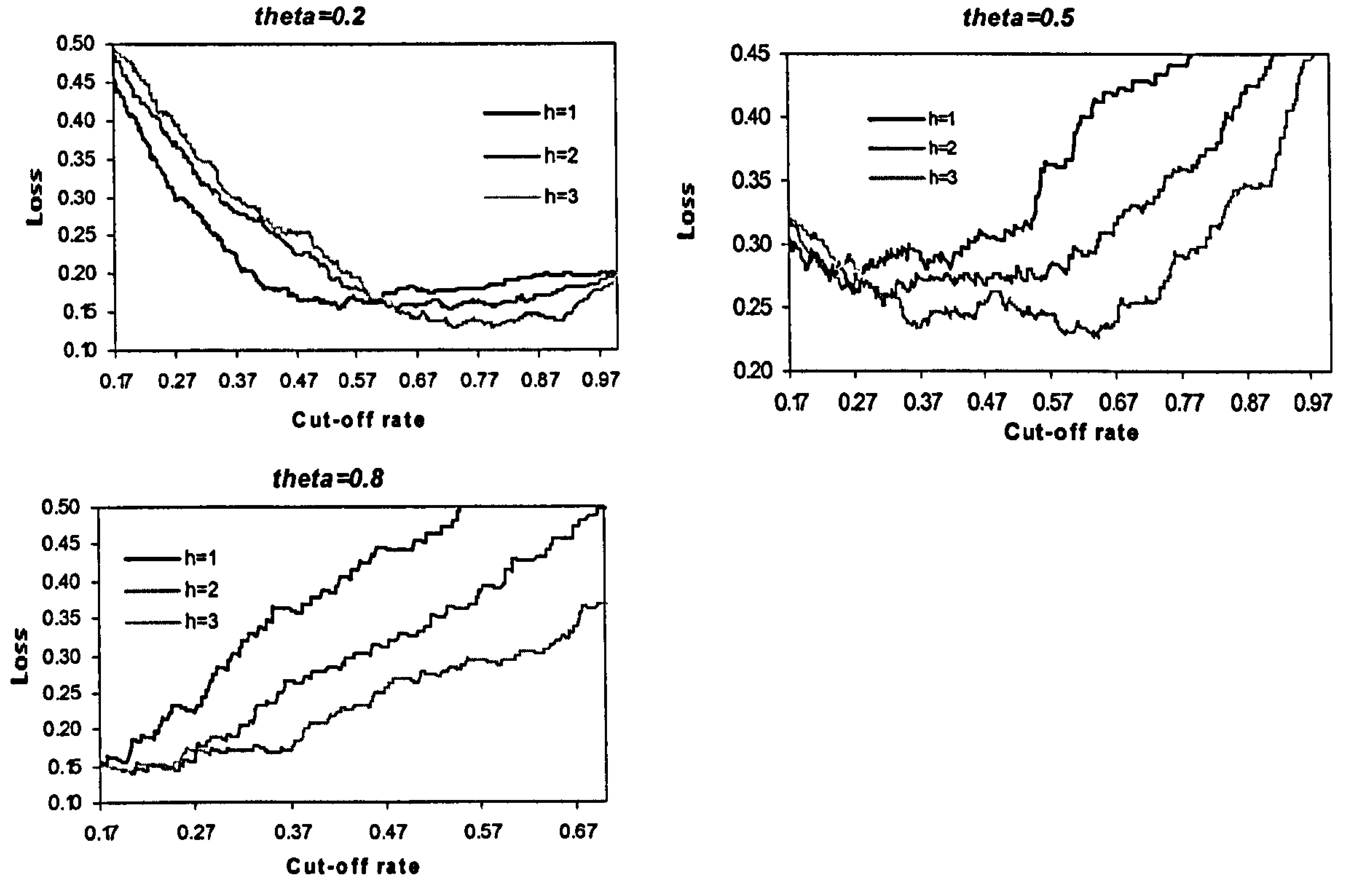


Figure 4.2: The Overall Loss for Different Cut-off Rate and Warning Horizon Combinations
(LOGIT-M)

Panel A: Investors Loss function



Panel B: Policymakers Loss function

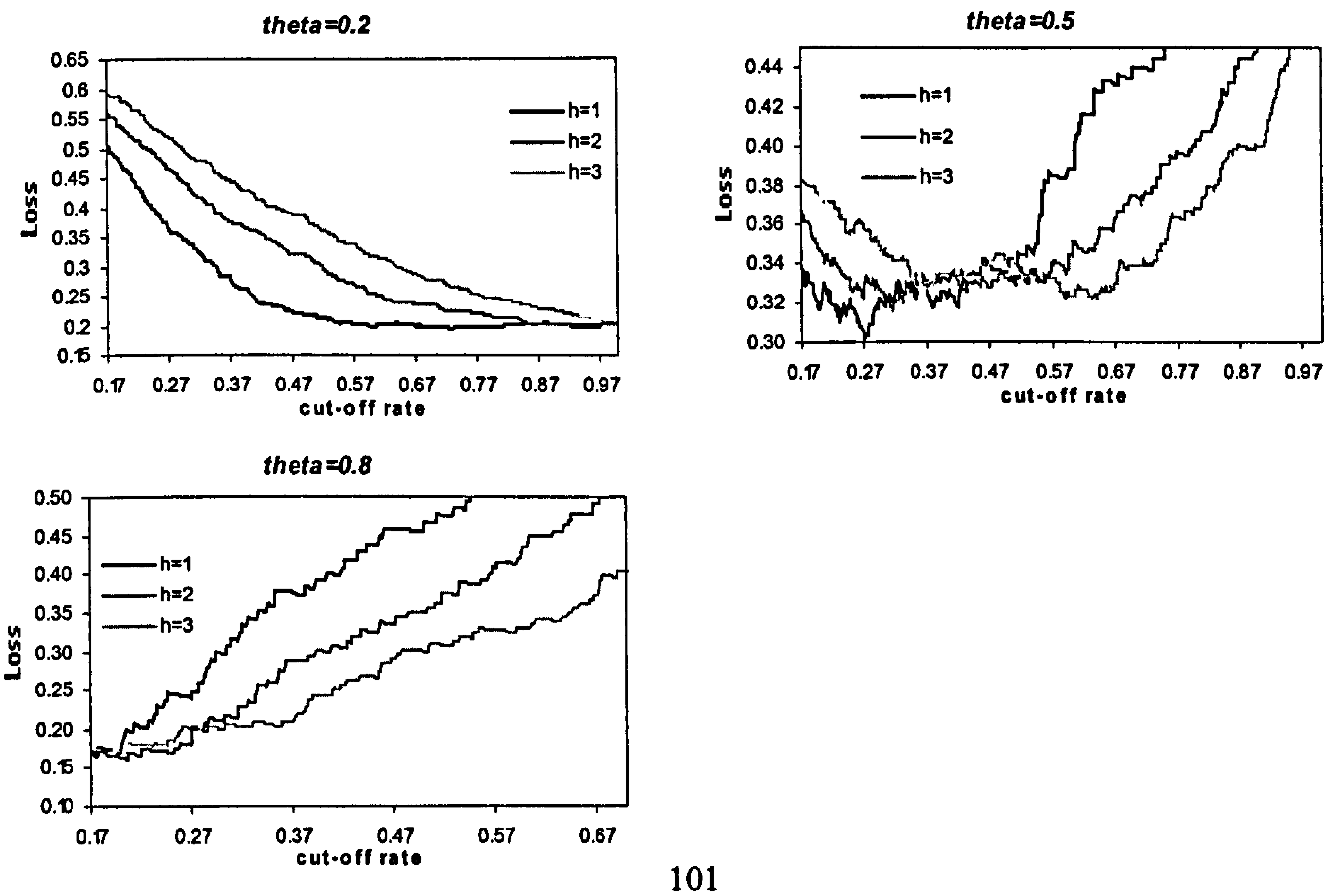
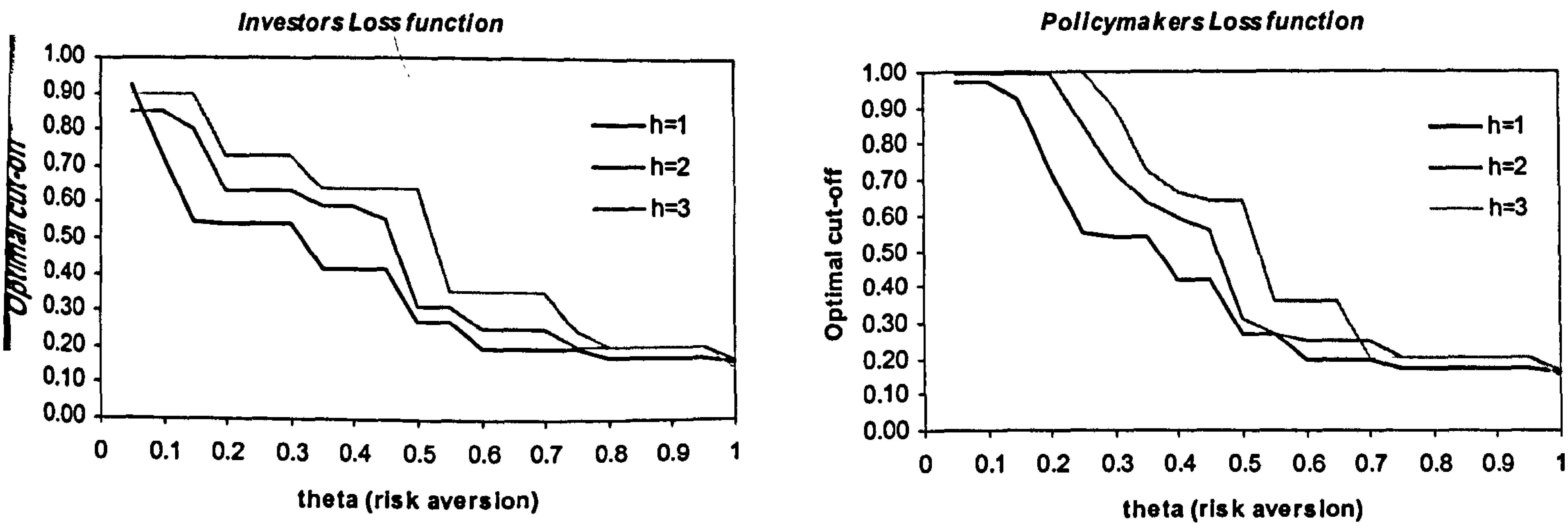


Figure 4.3: The Cut-off Rate and the Warning Horizon

Panel A: Optimal Cut-off Rate for Different Warning Horizons (LOGIT-M)



Panel B: The Overall Losses for Different Warning Horizons (LOGIT-M)

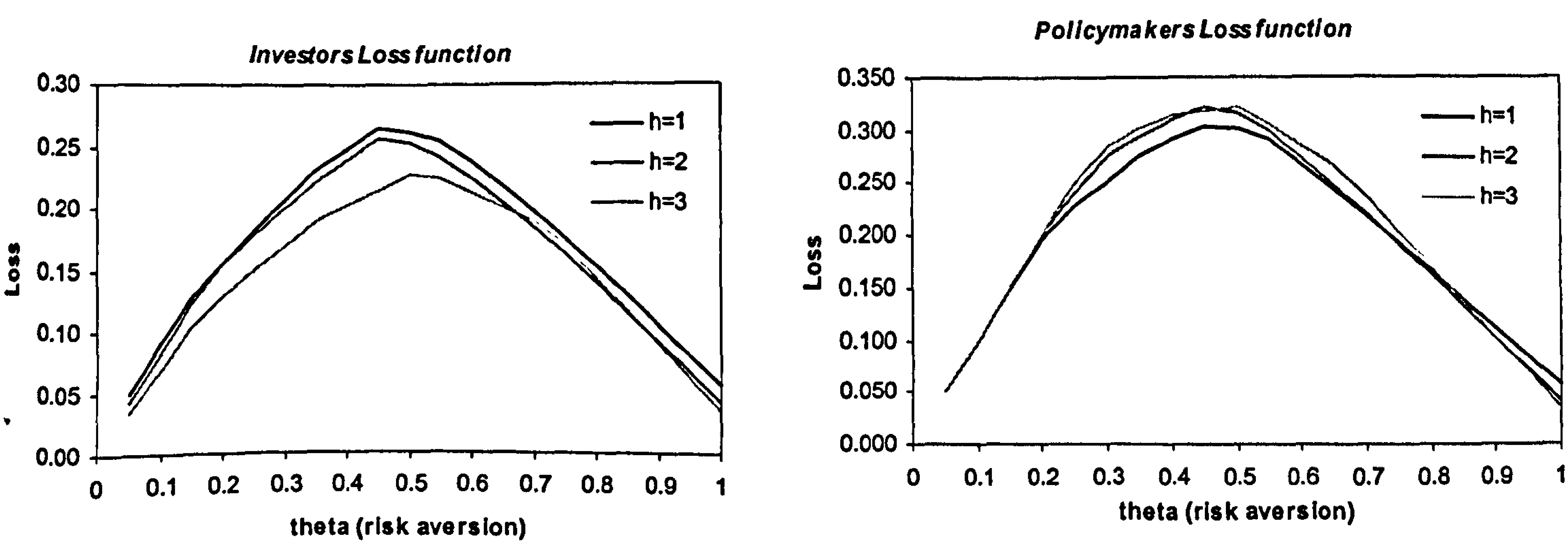
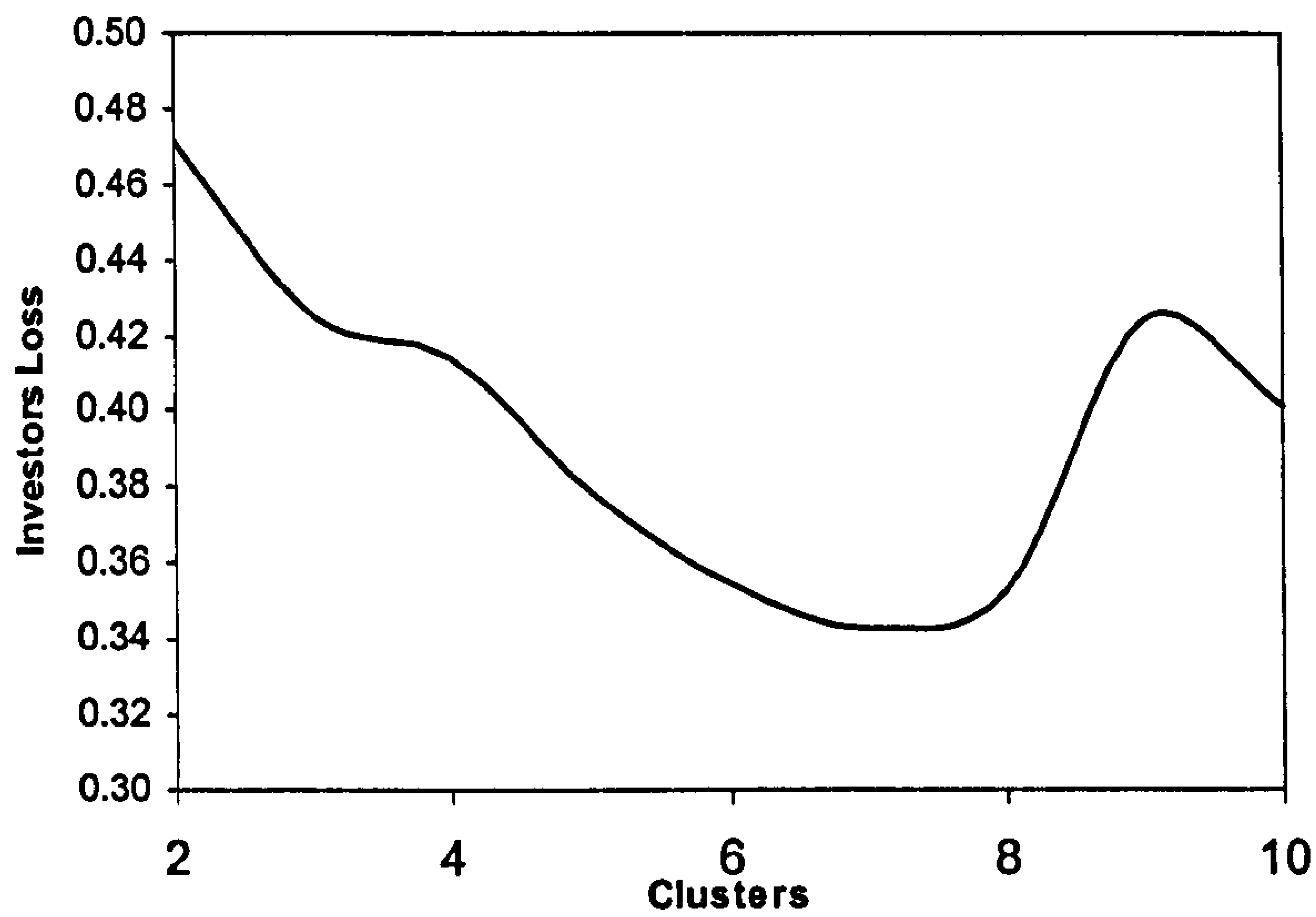


Figure 4.4: Calibration of the Number of Clusters

Panel A: Minimal Investors' Loss (θ=0.5)



Panel B: Optimal Number of Clusters for Each User

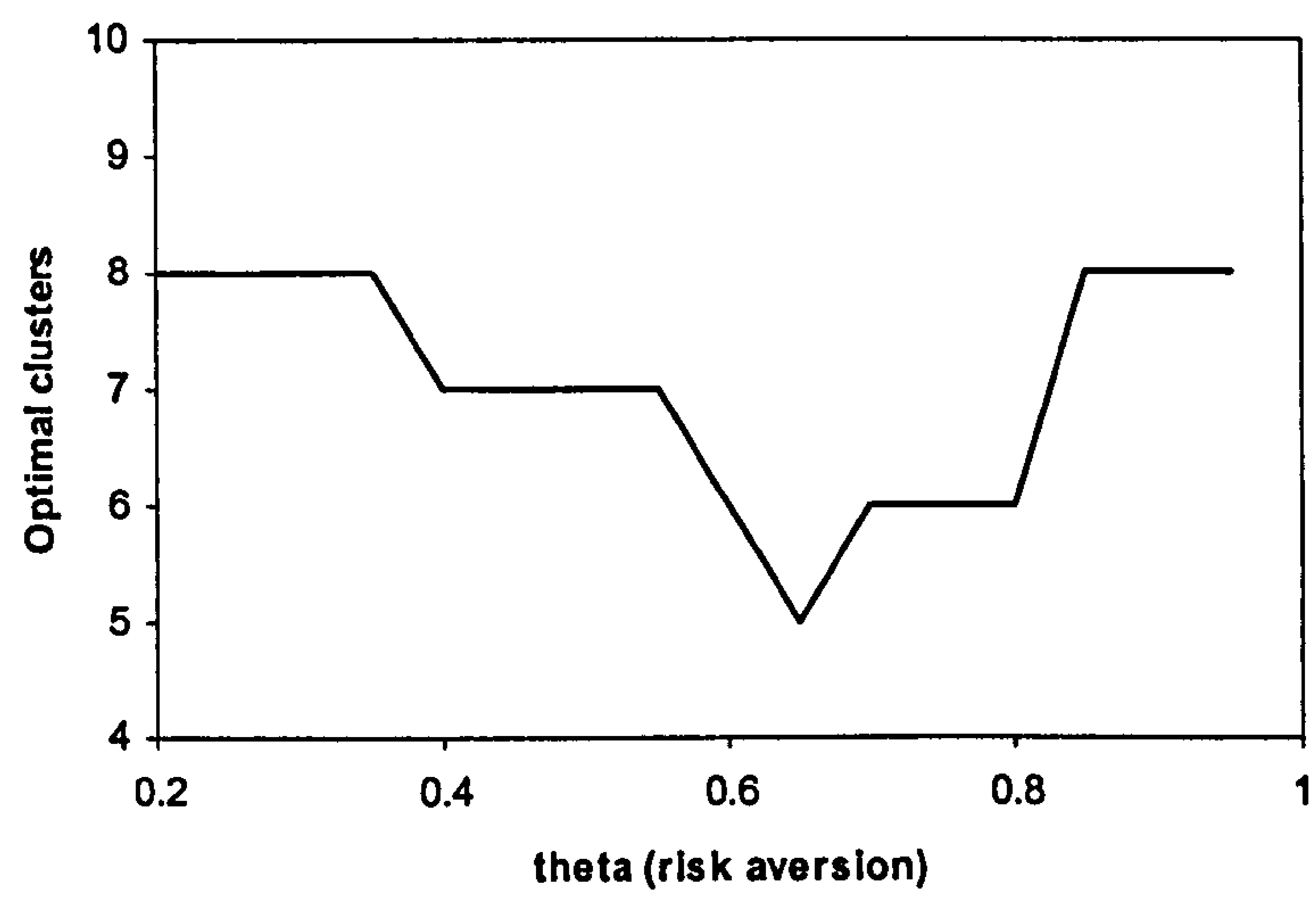
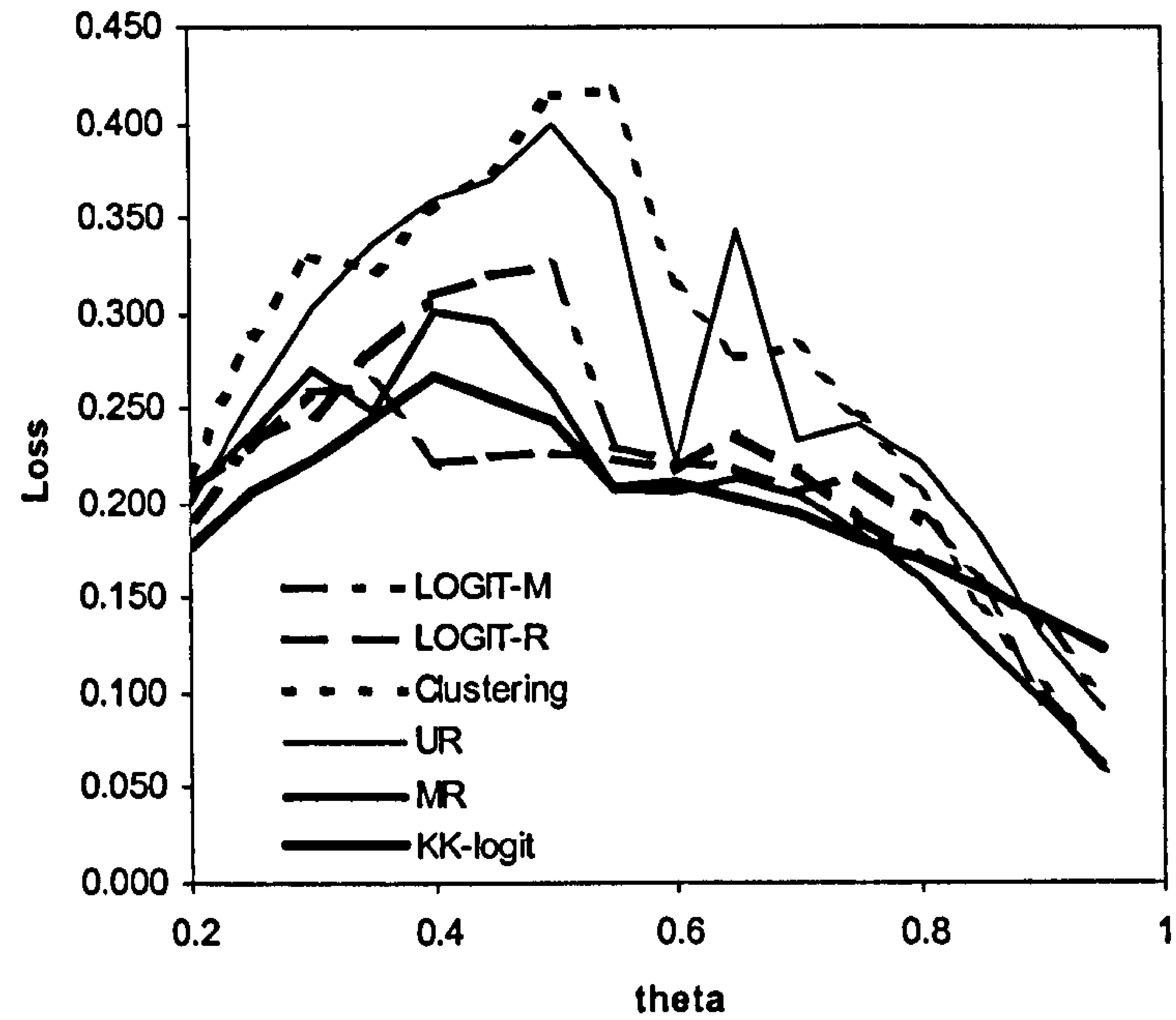
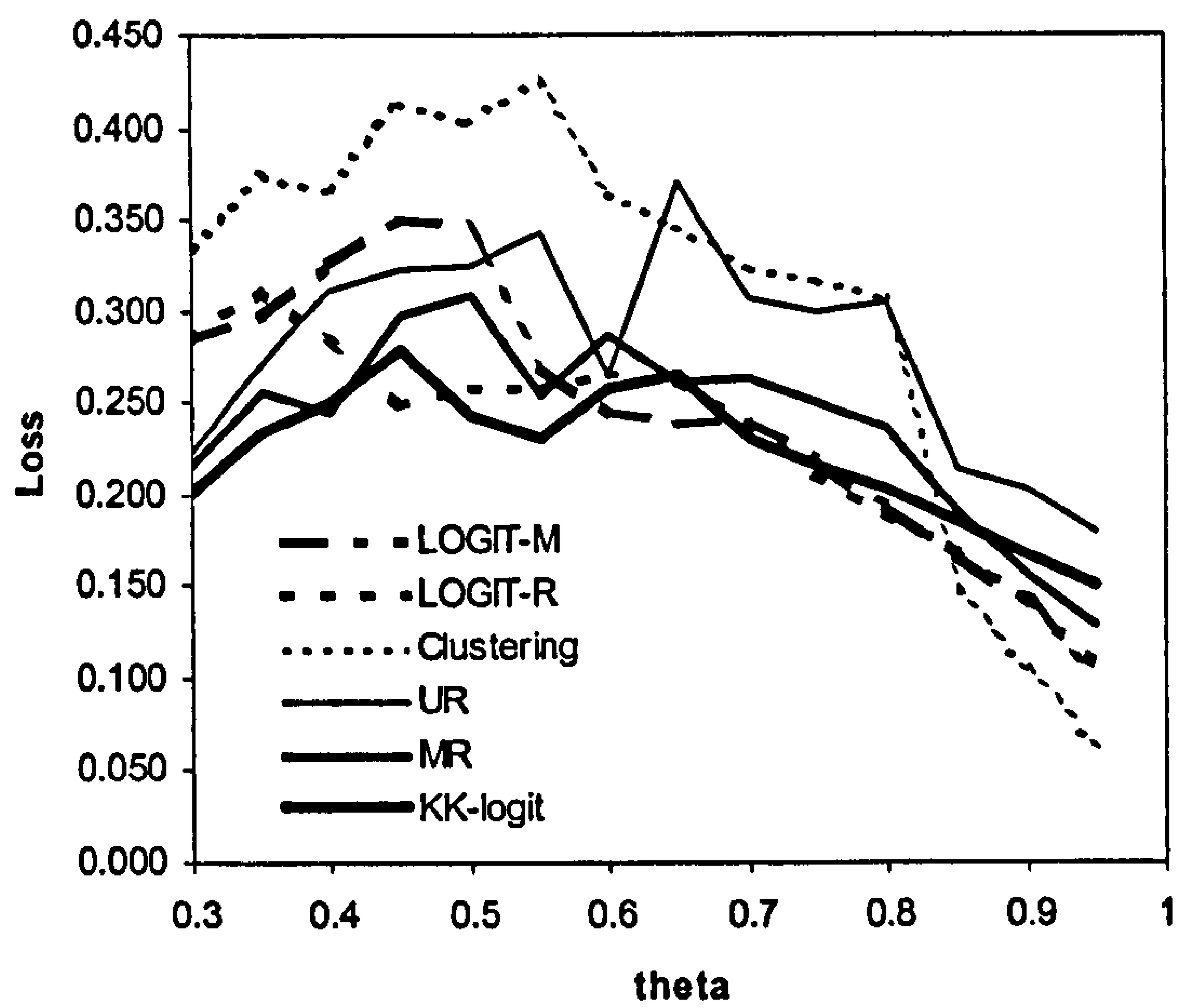


Figure 4.5: Overall Loss for Individual and Combined Forecasts

Panel A: Investors' Loss Function



Panel B: Policymakers' Loss Function



CHAPTER 5: A Comparative Analysis of Sovereign Credit Migration Estimators

5.1 Introduction

The development of credit risk management over the past few years has brought rating migration modeling to the forefront of attention. A chain of major sovereign defaults on foreign-currency denominated debts has been the main impetus for the growing interest — the 1994-1995 Mexican-peso crisis, the 1997 Asian currency crisis, the 1998 Russian ruble devaluation and, more recently, the 2001 credit failure in Argentina. These events increased the financial institutions concerns about their risk exposure to emerging market borrowers and their quest for better risk management procedures.

The fate of emerging markets with regard to international capital market access is inextricably linked to their external credit rating which plays a major role in modern credit risk management, valuation and international capital allocation. Financial institutions use credit ratings to feed their Value-at-Risk (VaR) analyses, to price credit risky loans, bonds and credit derivatives and to determine concentration limits. The Basel Committee on Banking Supervision, a regulatory body under the Bank of International Settlements, has proposed the New Basel Accord (Basel II, 2001) which permits banks to use internal credit rating systems to determine the regulatory capital against their credit exposure. Credit migration matrices are central inputs to industrial risk management tools such as J.P. Morgan's Credit Metrics, McKinsey's Credit Portfolio View and, more generally, VaR models for simulating the loss distribution of credit portfolios under extreme scenarios. For instance, the future value of a risky bank loan or bond is linked to its credit quality and thus, to the probability that it experiences an upgrade, downgrade or default.[1]

[1] The price of, say, a Baa bond has been found to increase by about 1-2% upon upgrade and to decrease by 30-50% upon default (Saunders and Allen, 2002, ch.6; Marrison, 2002, ch.18).

Furthermore, the need to insure against default losses has led to an ongoing expansion of credit derivative instruments. Prominent models for the term structure of credit-spreads (Jarrow et al., 1997) and the pricing of credit derivatives (Kijima and Komoribayashi, 1998) also build upon rating transition probabilities.

Risk management and capital allocation require a mapping of credit ratings into default probabilities and intra-rating transition (or migration) probabilities. A number of statistical issues undermine the reliability of conventional probability estimates in this context. One such problem is the relatively small sample of rating transitions available which becomes more pronounced when the focus is on specific categories of obligors such as financial institutions, European industrials or sovereigns.

The defaulted amount of sovereign debt and the subsequent scale of losses suffered in the last two and a half decades exceed by far those of corporate defaults. Nonetheless, most extant empirical studies mainly focus the latter. The application of existing corporate rating migration and credit risk management methods to sovereign credit ratings has received scant attention. Sovereign migration modeling is fraught with difficulties, mainly because of data limitations — sovereign rating histories (T) are relatively short and the cross-sectional dimension (N) is also small. The problem becomes even more acute for low credit-quality (emerging market) issuers because until very recently the ratings were mainly produced for industrialized sovereigns.[2] Rating transitions, particularly distant ones and defaults on sovereign bonds, are thus rare events.

The usual framework for modeling rating migration is a discrete Markov process which is time-homogenous. Hereafter we refer to the latter as just 'homogeneity' to simplify the exposition. Homogeneity means that the migration probabilities are constant over time. This assumption simplifies the estimation but its plausibility is questionable. In fact,

[2]In 1993 Moody's foreign currency bond ratings pertained to 24 sovereigns of which only 8 were non-industrialized countries. The number of rated sovereigns has increased substantially over the last decade (e.g. 72 in 2004) with a higher proportion of emerging economies (76% in 2004).

heterogeneity is a stylized fact in the context of corporate rating migrations.[3]

The first empirical analysis of rating migrations in a continuous Markov framework was conducted recently by Lando and Skodeberg (2002). Since then other studies have also applied continuous approaches, albeit exclusively in the area of corporate ratings. Discrete modeling entails loss of information regarding the exact timing of rating changes and the duration in each rating.

To our knowledge, only two studies focus on the estimation of sovereign transition matrices in the literature. Wei (2003) proposes a discrete multi-factor Markov chain model that accommodates heterogeneity while Hu et al. (2002) exploit bond ratings and sovereign default data in a discrete, homogeneous ordered probit framework. Very little is known on the finite-sample properties of credit migration matrix estimators particularly in the context of sovereigns. There are two comparative studies available (Jafry and Schuermann, 2004; Christensen et al., 2004) but they focus on corporate debt.

The purpose of this chapter is to compare different approaches to the estimation of rating transition probability matrices for sovereigns. Hereafter we often refer to the latter as just 'transition matrices' for expositional simplicity. In particular, we assess the bias and the sampling variability of rival estimators. The estimation methods we consider differ with respect to their assumptions — a discrete multinomial (also called cohort) method and two continuous hazard rate (also called duration) methods. The hazard methods differ in that one imposes the assumption of homogeneity in the underlying Markov process whereas the other relaxes it. A matrix-norm statistic based on the spectral decomposition is used to gauge the dynamics or overall mobility implied by the transition matrix.[4] In order to statistically compare the finite-sample properties of these rival estimators we deploy a bootstrap method that facilitates the empirical distribution of the transition

[3]The literature typically assumes also that the underlying Markov process is identical across issuers.

[4]Less dynamics means overall smaller probability mass at off-diagonal positions or equivalently, a smaller likelihood of transitions.

probabilities and of the mobility metric. We explore two conjectures. One is that there are efficiency gains from continuous (versus discrete) estimation methods in small samples of transitions. The other is that there is heterogeneity in the sovereign rating migration process which, if neglected, may lead to biased and inefficient estimates.

The literature on debt ratings is very extensive in the context of corporates but relatively limited for sovereigns. There is extensive evidence that sovereign and corporate credit ratings behave differently. This prompts the thought that competing estimation methods may have different properties in these two worlds. This provides a motivation for the present chapter in two directions. First, the properties of the underlying Markov process for corporates and sovereigns may be different. For instance, there is consensus that corporate rating transition probabilities are time-varying and therefore hazard rate methods that allow for this heterogeneity are superior. However, it is unclear whether this also applies to sovereigns. Second, given that sovereign ratings are remarkably more stable than corporate ratings, it is expected that continuous methods will bring gains relative to simple discrete ones in this context.

The contribution of this study to the sovereign credit risk literature is threefold. First, it investigates the potential value added of continuous estimators that incorporate full information on the exact timing of rating transitions and on rating duration. Second, it fills another gap by formally assessing for the first time whether imposing the assumption of homogeneity results in sovereign transition matrix estimators that exhibit more bias and sampling variability than heterogeneous estimators. Third, it provides the first tests for the plausibility of the homogeneous Markovian migration process for sovereigns. For instance, the presence of momentum and duration effects will invalidate it. For this purpose, we employ spectral analysis and panel logit models. We find that the continuous (homogeneous) hazard rate estimator produces more reliable default probability estimates than the discrete multinomial estimator. The efficiency of the former is further enhanced upon relaxing the homogeneity assumption. Relatively higher sovereign default probabilities are generally obtained from the homogenous estimator. There are significant downgrade momentum and duration effects, consistent with the extant evidence on corporate ratings.

The remainder of the chapter is organised as follows. Section 5.2 outlines the background literature. Section 5.3 introduces the Markov chain framework and discusses the three estimators under study. Section 5.4 introduces the dataset. Section 5.5 outlines the bootstrap simulation experiments. Section 5.6 discusses the results of the tests for non-Markov effects. A final section concludes.

5.2 Literature Review

The broad credit risk literature can be grouped in three strands. First, studies that model the rating migration process in order to derive transition matrix estimates include among others Altman and Kao (1992), Lucas and Lonski (1992), Carty and Fons (1993), Belkin et al. (1998), Duffee (1998), Helwege and Kleiman (1997), Bahar and Nagpal (2000) and Nickell et al. (2000). Second, studies that estimate risk-neutral credit migration matrices as inputs for pricing bonds and credit derivatives (Jarrow et al., 1997; Kijima and Komoribayashi, 1998; Bielecki and Rutkowski, 2000; Lando, 2000). The third strand focuses on the analysis of credit portfolio models, such as the CreditMetrics^{TM} (JP Morgan), the Credit Portfolio View (McKinsey) and the CreditRisk^{+} (CSFB), to simulate the distribution of credit assets and evaluate expected losses.[5] Our discussion below focuses primarily on those studies that directly investigate the characterisation of the rating migration with a few references to relevant work in the other two strands.

[5] For comprehensive surveys of credit portfolio models see Koyluoglu and Hickman (1998), Crouchy et al. (2000), Gordy (2000) and Saunders and Allen (2002).

5.2.1 Caveats in Rating Migration Analysis

The entry p_{ij} in a transition or migration matrix P gives the probability of being in the credit state j at time $t+1$, conditional upon being in state i at time t. Jarrow et al. (1997) introduce a distinction between implicit and explicit approaches to the estimation of migration matrices in the literature. The implicit approaches are based on market data (i.e. prices of risky zero-coupon bonds and assumptions on recovery rates) whereas the explicit approaches draw upon the rating and default history of obligors. Another distinction between the available estimators follows from their assumptions regarding the dynamics of the rating migration process. Regarding the explicit estimation, which is our focus in this study, different methods have been proposed ranging from simple discrete estimators to continuous ones that build on survival theory.

The transition matrix for Moody's seven broad ratings (Aaa, Aaa,..., C) plus the default state involves estimating 56 parameters if default is treated as an absorbing state. Accurate estimation presumes not only a sufficient number of observations but also enough cases in each rating as well as enough transitions. The use of a finer rating scale in order to increase the number of observed transitions has the shortcoming that more parameters need to be estimated and there are less issuers (transitions) per rating.[6] Insufficient data is a pervasive problem particulary for sovereign rating migrations. As in the case of corporates, defaults at the upper range of the rating spectrum are inexistent. Distant transitions are rare because migrations typically occur between neighboring ratings. Observations are also scarce at the lower end of the rating spectrum because the leading rating agencies have primarily focused on *investment grade* issuers.[7] Lastly, the number of sovereigns with rated bonds is relatively small (N) and the rating histories are short (T). All these issues raise concerns about the reliability of estimation and statistical inferences

[6]In 1982 Moody's expanded each coarse rating category into three sub-categories identified by three numbers, e.g. Aa1, Aa2, Aa3. The counterpart S&P sub-categorisation is AA+, AA, AA-.

[7]Issuers with no rating worse than Baa3 in the Moody's and BBB- in the S&P's systems, respectively.

on sovereign transition matrices.

The growing area of risk management has spurred interest on developing accurate sovereign credit migration estimators. One approach is to use samples of corporate bond ratings which have a much longer history going back to 1970.[8] However, this presumes that the sovereign migration process resembles that of corporates. But the empirical evidence reveals significant discrepancies. Jackson and Perraudin (2000) find that, on average, sovereign credit spreads are substantially lower than those of similarly rated corporates. Cantor and Packer (1996) find that sovereign ratings exhibit more discrepancies than corporate ratings across agencies. Nickell et al. (2000) show that the transition probabilities of US industrials differ significantly from those of similarly rated sovereigns.

Hu et al. (2002) seek to overcome the small-sample problem by augmenting the sovereign ratings available with default data on non-rated sovereigns from the UK Export Credit Guarantee Department (ECGD). As the latter incorporates defaults on commercial debt and trade credit obligations, a frequent event in the 1980s through to the early 1990s, this increases the sample size both in terms of countries and time span. The above information for rated and non-rated sovereigns is pooled and explained, through a (discrete) ordered probit model, in terms of macrovariables that proxy liquidity, solvency and economic conditions. The fitted rating histories are then used to estimate transition matrices which, they argue, can be combined with those from S&P's to increase accuracy.

By construction, the discrete multinomial approach typically adopted by leading rating agencies (Carty and Fons, 1993; Carty, 1997) and in most of the academic literature (Bangia et al., 2002) produces zero probability estimates for unobserved transitions. For corporate ratings, Lando and Skodeberg (2002) illustrate how continuous hazard rate methods can facilitate probability estimates of transitions (e.g. from Aa to default) that are not observed in the sample. Being able to estimate these probabilities is important because such transitions can actually occur given a large enough sample. This argument

[8]Moody's corporate ratings database contains about 60,000 annual observations from 1970 to today.

is implicitly supported by Basel II which establishes a minimum probability of 0.03% for such (unobserved) rare events. Further, a positive risk weight is assigned for sovereigns rated Aa to Baa which implies that Basel II recognizes the risk of default from these ratings.

5.2.2 Accuracy of the Transition Probability Estimates

Regarding the sampling variability or noise-to-signal of discrete migration probability matrix estimates, Bangia et al. (2002) find that the diagonal elements are quite accurate whereas the reliability of the off-diagonal entries falls as one moves away from the diagonal (i.e. transition between distant ratings). This result can be explained in terms of the persistence in credit ratings and the low frequency of distant transitions. Nickell et al. (2000) estimate standard errors for discrete transition probability estimates using a binomial distribution.[9] High standard errors are reported for low credit ratings because of the fewer observations per rating available (number of observations) and the greater volatility (probability of rating change) of rating transitions. Both studies provide binomial confidence sets for their probability estimates and conduct t-tests to assess the element-by-element differences between 'unconditional' and 'conditional' transition matrices. The latter condition on the rating history and/or the business cycle. The small sample alongside the discrete multinomial estimators used result in (meaningless) very wide confidence intervals for rare events. Using bootstrap confidence sets, Christensen et al. (2004) show that continuous transition matrix estimators are more accurate than the (industry standard) discrete multinomial estimators particularly for rare transitions from high ratings to default.

Lando and Skodeberg (2002) and Jafry and Schuermann (2004) deploy a continuous

[9]Each entry p_{ij} is cast as a binomial random variable — from rating i there will be a transition over a given horizon (e.g. one year) either to j or to $k = 1,..,K-1$, $k \neq j$ with probabilities p_{ij} and $(1-p_{ij})$, respectively. Assuming transition independence, the standard error for $\hat{p}_{ij}$ is the binomial standard deviation $\sqrt{\hat{p}_{ij}(1-\hat{p}_{ij})/N_i}$, where N_i is the number of sovereigns rated i at year-beginning.

heterogeneous hazard rate estimator for corporate ratings and find that it markedly differs from the discrete multinomial estimator but not so much from the homogeneous hazard rate estimator. However, they arbitrarily adopt a 1-year horizon in the analysis and acknowledge that over longer horizons the (time) heterogeneity will become more apparent and so their findings may not hold. This conjecture has not been explored as yet.

An interesting hypothesis is whether rival migration matrix estimators yield different portfolio loss distributions and hence, different VaR measures. To this end, Jafry and Schuermann (2004) test for the statistical and economic significance in the context of US corporate migration matrices. They corroborate that there are significant differences in the implied portfolio VaR capital from the discrete multinomial and the continuous hazard estimators. However, allowing for heterogeneity within the year appears less important, both statistically and economically, notwithstanding the evidence of heterogeneity in the rating evolution.

5.2.3 Non-Markov Effects and Time Variation in the Migration Process

Another issue that has spurred a considerable amount of research is the Markovian behaviour and time homogeneity of the migration process. In the celebrated Jarrow et al. (1997) framework, the dynamics of rating migration is modeled via a discrete, homogeneous Markov chain. The homogeneity assumption is critical as it rules out dependence on business cycles and non-Markov effects such as rating drift or momentum. Some attempts have been made recently at testing for non-Markov effects but only in the context of corporate ratings.

Momentum Effects

Several studies address the hypothesis of rating drift or momentum which implies that a previous upgrade/downgrade increases the probability of a future upgrade/downgrade.

Put differently, prior rating changes carry predictive power for the direction of future changes and so the current state does not fully determine the future transition iprobabilities. Downgrade and, to a very small extent, upgrade momentum in corporate ratings is supported by Lucas and Lonski (1992), Altman and Kao (1992), Carty and Fons (1993), Altman (1998), Kavvathas (2001) and Bangia et al. (2002) and Lando and Skodeberg (2002). The latter study utilizes a semiparametric exponential hazard rate model that allows for the momentum effect to vary across ratings. To have enough transition data for reliable inference they focus only on transitions to neighboring ratings.

All the above studies provide evidence of downgrade momentum. This is often attributed to the rating agencies' common practice of gradually downgrading an issuer (corporate or sovereign) rather than drastically reducing its rating.[10] Christensen et al. (2004) propose a continuous hazard rate model that accommodates downgrade momentum. They find that the default probability estimates for high credit ratings increase when accounting for momentum effects.

Duration Effects

The duration effect, another non-Markov property, is the relationship between the time spent in a given rating and the transition probability. This issue has been recently examined by Lando and Skodeberg (2002) and Kavvathas (2001) and both confirm the earlier findings by Carty and Fons (1993) that uncover duration effects in corporate ratings, although the evidence on the direction of the relationship is conflicting. Lando and Skodeberg (2002) find negative duration dependence which means that the longer an issuer stays in a rating the lower is the probability of a rating transition. The duration effect is documented for all ratings and both for upgrades and downgrades. Kavvathas (2001) finds evidence of positive duration dependence instead. One explanation for this

[10]This is linked to morel hazard (Sy, 2002) because of the the client-provider relationship between issuers and rating agencies (Altman and Saunders, 1998). The latter receive considerable fees from the issuers for rating them. A severe downgrade will leave the customers unhappy.

conflicting evidence is that, although both studies use continuous hazard rate methods, the distributions (exponential and Weibull, respectively) they assume for the hazard rate can only capture monotonic duration effects. Du (2003) argues that the duration dependence in corporate rating migrations is not monotonic but Λ- and U-shaped for upgrades and downgrades, respectively. In particular, he empirically shows that the upgrade probability initially increases with duration, peaking after about two years and then decreases. The opposite pattern is observed for the downgrade probability.

A rationale for the decline in downgrade probability with duration is again the tendency of rating agencies to downgrade an obligor notch by notch until their 'target' rating is reached. Thus one observes (sample) low ratings with very short durations but high downgrade risk and vice versa for high ratings.[11] On the other hand, upgrades are positively linked to duration because agencies are more willing to drastically upgrade an obligor once it has spent a considerable amount of time at the current rating.

Business Cycle Effects

A few studies address the issue of business cycle time-heterogeneity in corporate rating migrations. Lucas and Lonski (1992) were the first to provide empirical evidence of these cyclical effects which are particularly present in non-investment grade (or speculative) bonds. Subsequently, this has been corroborated by Helwege and Kleimen (1997) by documenting significant correlation between corporate default probabilities and the state of the economy. Belkin et al. (1998), Nickell et al.(2000), Bangia et al. (2002) and Kavvathas (2001) explicitly model this cyclical heterogeneity using dummies for the state of the business cycle and exogenous macrovariables such as interest rates, stock return volatility and stock returns. Nickell et al. (2000) document that the impact of business cycles on the transition probabilities differs between banks, industrials and other types

[11] Du (2003) provides evidence that the duration effect in downgrades loses significance when momentum is controlled for.

of borrowers (e.g. financials, insurance or sovereigns) as well as across US and non-US investment grade issuers. Overall there is evidence that business cycle variations have a strong impact on transition probabilities and especially on default probabilities of non-investment grade issuers. In this light, Lando (2000) and others allow for business cycle effects in bond pricing. On the other hand, Blume et al. (1998) and Du (2003) show that the declining trend in ratings over the recent decade is linked to more stringent standards (rating agencies) rather than to the business cycle.

Kim (1999) and Nickell et al. (2000) propose an ordered probit model for rating migrations. In this context, given an initial rating the probability that an issuer migrates to another rating is linked to specific latent factors via the cumulative normal distribution. Nickell et al. (2000) attempt to disentangle the effects of borrower characteristics (industry and country) and business cycle fluctuations on the rating transition probabilities through the use of dummy variables. They then assume that business cycle evolution is driven by a stochastic three state (peak, trough and normal) Markov process. Between-rating variation is analysed by running different ordered probit models for sub-samples of issuer years starting with the same rating.

Wei (2003) proposes a multi-factor Markov chain model for rating migrations. The model facilitates the estimation of transition matrices which vary over time according to rating-specific latent variables that may reflect business or credit cycle fundamentals. A key aspect of the model is that it allows the transition probabilities for each rating to respond differently to external factors. While Wei advocates the application of this (highly parametrised) model for both corporate and sovereign debts, he recognizes the limitations in the later case due to the small samples available.

5.3 Methodology

This section discusses three distinct credit migration estimation methods. We start by outlining the homogeneous Markov chain model which is the assumed baseline data generating process (DGP) for credit ratings.

5.3.1 The Time-homogeneous Continuous Markov Chain Model

The literature on migration matrix estimation essentially rests on two assumptions. First, the future rating is independent of the rating history (first-order Markov property). Second, the transition probabilities remain constant over time (homogeneity).

Let S denote the transition space and $i = 1, 2, ..., K$ the available credit ratings. Moody's bond rating system has seven coarse states (Aaa, Aa, A, Baa, Ba, B, C), twenty three finer states (Aaa1, Aaa2, Aaa3, Aa1,..., Ca) plus the default (D) state. This study adopts the broad credit rating scale so $i = 1$ is the highest (Aaa) rating and $i = K = 8$ denotes default. Let $P(s, t)$ denote the $K \times K$ transition probability matrix generated by a continuous Markov chain η so that

$$p_{ij}(s,t) = \Pr(\eta_t = j \,|\eta_s = i) \ , \ s < t \tag{5.1}$$

is the probability that a sovereign rated i at time s migrates to rating j at time t. At this point it is important to recall the first-order Markov property formally,

$$\Pr(\eta_t = j \,|\eta_0 = i_0, \eta_1 = i_1, \eta_2 = i_2, ..., \eta_s = i) = \Pr(\eta_t = j \,|\eta_s = i) \text{ for } 0 < 1 < ... < s$$

which further implies that

$$P(s,t) = P(s,u)P(u,t), \ s < u < t \tag{5.2}$$

If the Markov chain is *homogeneous,* then the transition matrix only depends on the transition horizon ($\Delta t \equiv t - s > 0$) but not explicitly on time

$$P(t-s) = P(t-u)P(u-s), \ s < u < t \tag{5.3}$$

and thus we have a family of transition matrices, $P_{\Delta t}$, indexed by Δt. Equation (5.3) further implies that the transition matrix for horizon $n\Delta t$ is simply $P^n_{\Delta t}$.[12] Hereafter, P denotes the transition matrix over a one-period (i.e. $\Delta t = 1$) horizon

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1K} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2K} \\ \vdots & & \ddots & & \vdots \\ p_{K-1,1} & p_{K-1,2} & p_{K-1,3} & & p_{K-1,K} \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{5.4}$$

where $p_{ij} \geq 0 \; \forall \; i, j$, $\sum_{j=1}^{K} p_{ij} = 1 \; \forall \; i$ and default is treated as an absorbing state.

5.3.2 Discrete Multinomial (DM) Estimator

The conventional transition matrix estimator is a discrete multinomial (DM) approach based on annual migration frequencies. Transitions away from a given state $i = 1, 2, ..., K$ over a one-year horizon are assumed to follow a multinomial distribution with $K - 1$ outcomes and associated probabilities p_{ij}, $j = 1, .., K$, $i \neq j$. Let $N_i(t)$ denote the number of sovereigns that start year t in state i and $N_{ij}(t, t+1)$ the number of sovereigns that migrate to j by the start of year $t+1$. The migration frequency over $[t, t+1]$ is $\frac{N_{ij}(t,t+1)}{N_i(t)}$.

Let us assume homogeneous Markov chain dynamics for the rating process (time independence) and independence between the multinomial experiments for different issuers (cross-section independence). Accordingly, the maximum likelihood estimator (MLE) of the one-year transition probability based on the pooled ratings is defined as

$$\hat{p}_{ij} = \sum_{t=1}^{T} w_i(t) \frac{N_{ij}(t, t+1)}{N_i(t)} = \frac{\sum_{t=1}^{T} N_{ij}(t, t+1)}{\sum_{t=1}^{T} N_i(t)} \tag{5.5}$$

where T is the number of sample years and $w_i(t) = \frac{N_i(t)}{\sum_{t=1}^{T} N_i(t)}$ is the weight for the year t migration frequency. Many studies use instead the sample average of the year-on-year migration frequencies $\hat{p}_{ij} = \frac{1}{T} \sum_t \frac{N_{ij}(t,t+1)}{N_i(t)}$ as estimator (Bangia et al., 2002; Hu et al., 2002). However, this coincides with (5.5) only in the special case where the number of

[12] See Norris (1997) for a depth-in discussion of continuous Markov chains.

sovereigns rated i remains constant over the sample period, $N_i(t) = N_i, t = 1, 2, ..., T$. This implies that the annual inflow is equal to the outflow which is implausible.

Israel et al. (2001) argue that transition probability matrices obtained from the discrete estimator (5.5) are typically not consistent with a continuous Markov chain process mostly because of the concentration of probability mass around the main diagonal. Moreover, the one-year discrete estimator (5.5) neglects information about within-year rating transitions and about rating duration or the time spent by an issuer in each rating. In the context of sovereigns, where the available rating histories are short (T) and over small cross-sections (N), it is crucial to account for as much additional information as possible. This motivates the following estimators.

5.3.3 Continuous Hazard Rate Estimators

We outline below the relevant survival theory concepts in a general framework where the transition probabilities are time-dependent or heterogeneous. Next we discuss a hazard rate estimator that assumes homogeneity and a more general heterogeneous estimator.

Background on Survival Analysis

Let the random variable τ_i denote the duration of a sovereign in rating i or 'survival' time. Let Λ_t be the heterogenous Markov chain generator or intensity matrix

$$\Lambda_t = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & \dots & \lambda_{1K} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \dots & \lambda_{2K} \\ \vdots & \vdots & \ddots & & \vdots \\ \lambda_{K-1,1} & \lambda_{K-1,2} & \lambda_{K-1,3} & & \lambda_{K-1,K} \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

where the off-diagonal transition intensity

$$\lambda_{ij}(t) \equiv \lim_{\Delta t \to 0^+} \frac{p_{ij}(t \leq \tau_i \leq t + \Delta t \mid \tau_i \geq t)}{\Delta t} \geq 0 \tag{5.6}$$

is defined as the instantaneous rate of transition from rating i to j at $\tau = t$ conditional upon being in state i ('survival') up to time t. The diagonal element

$$\lambda_{ii}(t) \equiv \lambda_i(t) \equiv - \sum_{j=1, j\neq i}^{K} \lambda_{ij}(t), \ i \neq j \tag{5.7}$$

is the *hazard rate function* or the instantaneous rate of transition from i at time t conditional upon being in state i up to time t.. The probability of leaving rating i over any time horizon Δt can be approximated by $\lambda_i(t)\Delta t$.

The behavior of τ_i can be characterized through the hazard rate function, the survivor function or the probability density function. The *survivor function* is defined as

$$F_i(t) = \Pr(\tau_i \geq t) = \Pr(\eta_u = i \ \forall \ u \in (s,t] \, | \eta_s = i) \, , \ 0 < t < \infty \tag{5.8}$$

and it represents the probability that the process does not leave state i before time t.[13] The *probability density function* (pdf)

$$f_i(t) = \lim_{\Delta t \to 0^+} \frac{\Pr(t \leq \tau_i \leq t + \Delta t)}{\Delta t} \tag{5.9}$$

gives the instantaneous rate of transition from state i at time t. Combining equations (5.6)-(5.9) it follows that the three specifications are interrelated via

$$\lambda_i(t) = \frac{f_i(t)}{F_i(t)} = \frac{-d \log F_i(t)}{dt} \tag{5.10}$$

so that by integration and with the initial condition $F(0) = 1$ we have

$$F_i(t) = \exp(- \int_0^t \lambda_i(u) du) \tag{5.11}$$

Using equation (5.10) the pdf of τ_i can be written as

$$f_i(t) = \lambda_i(t) \exp(- \int_0^t \lambda_i(u) du) \tag{5.12}$$

[13]This probability is different from $p_{ii}(s,t)$ in (5.1). The latter gives the probability of being in state i both at times s and t but the process could have been in any other state after s and before t.

Exponential and Weibull densities are often employed (for the hazard rate function λ_i) to characterize failure time.[14] The former assumes a constant hazard rate, that is, the instantaneous migration rate from i is homogeneous or $\lambda_i(t) = \lambda_i$. The exponential survivor and density functions are

$$F_i(t) = \exp(-\lambda_i t) \quad \text{and} \quad f_i(t) = \lambda_i \exp(-\lambda_i t),$$

respectively. The two-parameter Weibull distribution has the hazard rate function

$$\lambda_i(t) = \lambda_i \beta (\lambda_i t)^{\beta - 1}, \ \lambda_i, \beta > 0 \tag{5.13}$$

which allows for explicit heterogeneity (time-dependence) through a power function specification. The shape parameter β dictates the type of time-dependence, i.e. whether failure intensity increases or decreases with the duration of the current state. The hazard function is monotonically increasing for $\beta > 1$ (positive duration dependence), decreasing for $\beta < 1$ (negative duration dependence) and equal to the constant hazard for $\beta = 1$ (no duration dependence). The Weibull survivor and density functions are thus

$$F_i(t) = \exp\{-(\lambda_i t)^\beta\} \quad \text{and} \quad f_i(t) = \lambda_i \beta (\lambda_i t)^{\beta-1} \exp\{-(\lambda_i t)^\beta\},$$

respectively. The explicit heterogeneity imposed in this formulation breaks down the Markov property. A constant hazard rate, as in the exponential distribution or Weibull for $\beta = 1$, is sufficient for the Markov property to hold and necessary to have an homogeneous Markov process.

Homogeneous Hazard Rate (HHR) Estimator

The entry (i, j) of the transition intensity matrix Λ is driven by a random duration τ_i which is subject to censoring. Censoring refers to the fact that durations are discontinued at both ends of the sample. As one does not know for how long outside the sample window

[14] A detailed exposition of failure time analysis can be found in Kalbfleisch and Prentice (2002).

each individual survived in the current rating, we have a case of right censoring.[15] Rating withdrawals also result in right censored durations.[16]

In the case of credit rating transitions, there are $K-1$ outcomes (K possible ratings including default). A likelihood function is formulated under the assumption of independence across durations. The maximisation approach to estimate Λ accounts for the contribution of each possible transition or, equivalently, duration (i.e. from i to j, $j=1,...K$, $j \neq i$) in the overall transition probability for rating i. In this framework, the contribution of duration t_1 that involved a transition from state i to state j is

$$f_{ij}(t) = \lambda_{ij}(t)\exp(-\int_0^{t_1}\left\{\sum_{j=1,j\neq i}^{K}\lambda_{ij}(t)\right\}du \tag{5.14}$$

and for N_{ij} migrations occurring at times t_m, $m=1,...,N_{ij}$, and N_{ic} censored durations (at rating i) occurring at t_m, $m=1,...,N_{ic}$, the likelihood function for the transition from state i is formulated as

$$L_i = \left[\prod_{j=1,j\neq i}^{K}\prod_{m=1}^{N_{ij}}\lambda_{ij}(t)\exp(-\int_0^{t_m}\left\{\sum_{j=1,j\neq i}^{K}\lambda_{ij}(t)\right\}du)\right]\prod_{m=1}^{N_{ic}}\exp(-\int_0^{t_m}\left\{\sum_{j=1,j\neq i}^{K}\lambda_{ij}(t)\right\}du)$$

The third product above is the contribution of the right censored observations to the likelihood. The log-likelihood function is

$$\log L_i = \sum_{j=1,j\neq i}^{K}\left[N_{ij}\log(\lambda_{ij}(t)) - \sum_{m=1}^{N_{ij}}\int_0^{t_m}\left\{\sum_{j=1,j\neq i}^{K}\lambda_{ij}(t)\right\}du\right] - \sum_{m=1}^{N_{ic}}\int_0^{t_m}\left\{\sum_{j=1,j\neq i}^{K}\lambda_{ij}(t)\right\}du$$

If the intensities are homogenous, $\Lambda_t = \Lambda$, maximisation of the log-likelihood gives a simple closed-form solution for the transition intensities

$$\frac{\partial \log L_i}{\partial \lambda_{ij}} = \frac{N_{ij}}{\lambda_{ij}} - \sum_{m=1}^{N_{ij}} t_m - \sum_{m=1}^{N_{ic}} t_m = 0 \Longrightarrow \hat{\lambda}_{ij} = \frac{N_{ij}}{\sum_{m=1}^{N_i} t_m} \tag{5.15}$$

where $N_{ij}(T)$ is the number of transitions from i to j, ($i \neq j$), within the period $[0,T]$ such that $N_i = N_{ij} + N_{ic}$. The above MLE can also be written as $\hat{\lambda}_{ij} = \frac{N_{ij}(T)}{\int_0^T Y_i(u)du}$ where

[15] Left censoring refers to the corresponcing situation at the beginning of the sample window.

[16] A withdrawn rating (WR in Moody's jargon) indicates the interruption of credit quality assessment of a particular issuer due to non-credit related factors, for instance, maturing of outstanding debt.

$Y_i(u)$ is the number of sovereigns rated i at time u and so $\int_0^T Y_i(u)du$ gives the total time spent in state i by all the sovereigns in the sample. The transition probability matrix estimate over a Δt horizon is

$$\hat{P} = \exp(\hat{\Lambda}\Delta t), \ \Delta t \geq 0 \tag{5.16}$$

where $\exp(\hat{\Lambda}\Delta t) = \sum_{k=0}^{\infty} \frac{(\hat{\Lambda}\Delta t)^k}{k!}$. So under homogeneity, a simple mapping from transition intensities to transition probabilities exists.

The homogeneous hazard rate (HHR) estimator based on a continuous transition intensity matrix, equation (5.16), is expected to be more efficient than the DM counterpart, (5.5), for several reasons. First, over a given time horizon of one year (Δt), if a country migrated from Aaa$\longrightarrow$Aa$\longrightarrow$ A, then the transition to the intermediate state Aa contributes to the estimation of the transition probability from Aaa to Aa through (5.15). In the discrete framework, however, this information is ignored. Second, the continuous estimator accounts for the exact date when a sovereign receives a new rating and also for each rating duration. In the above example, the time spent in the intermediate state Aa is accounted for in the estimation of the transition intensity for rating Aa. Third, it readily accommodates right censoring by using information for the obligors up to the day of withdrawal and thus obligors ending the year in the WR status will not be discarded as in the DM estimator in (5.5). Fourth, it generally yields non-zero probabilities of rare transitions between states, even if they are not observed in the sample, as long as an indirect transition from one state to the other occurs. The DM estimates for these unobserved direct transitions would have been zero. For instance, suppose that no direct defaults from state Aa are observed, but we observe migrations from Aa to B and then to default. Then as long as the probability of migrating to B and the probability of default from B are both positive, the continuous estimate for the default probability from Aa is strictly positive. Finally, the estimates of the continuous transition intensity matrix can be easily transformed into transition probabilities over any time horizon.

Non-homogeneous Hazard Rate (NHHR) Estimator

The nonparametric approach of Aalen and Johansen (1978) facilitates a generalization of the above continuous hazard rate estimator to allow for heterogeneity in the underlying Markov process. Let $P(s,t)$ be the rating transition matrix over the horizon $[s,t]$, as defined in (5.1). In the heterogeneous setting, intensities and transition probabilities are linked by means of the *cumulative intensity* function

$$A_{ij}(t) = \int_0^t \lambda_{ij}(u)du, \ A_{ii}(t) = -\sum_{j \neq i} A_{ij}(t)$$

defined in Gill and Johansen (1990). The transition matrix is given by

$$P(s,t) = \prod_{[s,t]}(I + dA) \equiv \lim_{\max|t_m - t_{m-1}| \to 0} \prod_m (I + A(t_m) - A(t_{m-1})) \tag{5.17}$$

where $s \leq t_1 \leq t_m \leq t$. Assuming n transitions within the $[s,t]$ horizon, $P(s,t)$ can be consistently estimated by means of the Aalen-Johansen non-parametric product limit

$$\hat{P}(s,t) = \prod_{i=1}^{n}[I + \Delta\hat{A}(T_i)] \tag{5.18}$$

where T_i denotes a point in time over $[s,t]$ where a transition occurred, n is the total number of transitions (i.e. for an annual horizon, the number of days in a year where at least one transition occurs) and

$$\Delta\hat{A}(T_i) = \begin{bmatrix} -\frac{\Delta N_{1.}(T_i)}{Y_1(T_i)} & \frac{\Delta N_{12}(T_i)}{Y_1(T_i)} & \frac{\Delta N_{13}(T_i)}{Y_1(T_i)} & \cdots & \frac{\Delta N_{1K}(T_i)}{Y_1(T_i)} \\ \frac{\Delta N_{21}(T_i)}{Y_2(T_i)} & -\frac{\Delta N_{2.}(T_i)}{Y_2(T_i)} & \frac{\Delta N_{23}(T_i)}{Y_2(T_i)} & \cdots & \frac{\Delta N_{2K}(T_i)}{Y_2(T_i)} \\ \vdots & \vdots & \ddots & & \vdots \\ \frac{\Delta N_{K-1,1}(T_i)}{Y_{K-1}(T_i)} & \frac{\Delta N_{K-1,2}(T_i)}{Y_{K-1}(T_i)} & \cdots & -\frac{\Delta N_{K-1.}(T_i)}{Y_{K-1}(T_i)} & \frac{\Delta N_{K-1,K}(T_i)}{Y_{K-1}(T_i)} \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix} \tag{5.19}$$

where $\Delta N_{lj}(T_i)$ is the number of transitions from state l to j at time T_i.[17] The diagonal elements $\Delta N_{l.}(T_i)$ are the number of transitions away from state l at time T_i so that the

[17] $N_{lj}(t,t+1)$ counts the total number of transitions observed from l to j from the starting date t until $t+1$ and $\Delta N_{lj}(T_i)$ is an increment of this process.

sum of each row is zero and the rows of $I+\Delta\hat{A}(T_i)$ sum to 1. The number of sovereigns in state l just before time T_i is denoted $Y_l(T_i)$. The off-diagonal entries $\{\Delta\hat{A}(T_i)\}_{lj}, l \neq j$ thus represent the fraction of sovereigns at state l just before time T_i that migrated to state j at time T_i. The Aalen-Johansen method can be seen as the DM estimator (5.5) extended to infinitely short-time intervals (i.e. time points T_i). Finally, the non-homogeneous hazard rate (NHHR) estimator of the transition matrix is

$$\hat{P} = w_0\hat{P}(t_0, t_1) + w_1\hat{P}(t_1, t_2) + ... + w_{T-1}\hat{P}(T-1, T) \tag{5.20}$$

where $\hat{P}(t_0, t_1)$, $\hat{P}(t_1, t_2)$ and so forth are particularizations of (5.18) for sequential, non-overlapping intervals and w_i is the proportion of rated issuers at t_i.

5.3.4 Transition Horizon

The discrete estimation of migration matrices is inextricably linked to the notion of transition horizon. The latter can be simply defined as the regular time interval at which the credit ratings, and thus rating transitions, are observed. For instance, for a quarterly transition matrix the discrete estimator will capture transitions that occur between the beginning and the end of quarterly time periods. Since sovereign credit quality is continuously reviewed by rating agencies, DM estimators based on shorter transition horizons can reflect better the dynamics of the rating process. However, shorter horizons have the shortcoming that transitions between distant ratings are unlikely to be observed. Continuous estimation methods exploit the observed rating histories (without having to specify a fixed transition horizon) to provide transition intensities at any point in time.

The appropriate transition horizon in a discrete framework is dictated by the application. For credit portfolio analysis, one year or longer is standard practise. Shorter horizons are useful for credit-risky bond and credit derivative pricing models where cash flows occurring at all dates need to be weighted accordingly by the corresponding survival probability. The discrete transition matrices published by rating agencies are based on a

1-5 year horizon. Longer horizons have the property that they smooth out noisy data.

In our empirical analysis below the DM estimator is based on non-overlapping h-year horizons with $h \in \{1,2,3\}$. To illustrate, for $h = 2$, using the observed ratings (over 23 years) we compute two-year transition matrices by pairing the years and recording transitions in each of the 11 biannual periods. This approach has to discard the first year of data, 1981. For the three-year matrix, we leave out the first two years 1981-1982 and use information based on 7 triennial non-overlapping periods. In some studies they use overlapping horizons. For example, S&P's and Moody's estimate two-year transition matrices by counting the biannual transitions, $N_{ij}(t, t+2)$, sequentially for every year in the sample (see Bangia et al., 2002). For a sample covering 22 years, this implies using twice as many observations as with non-overlapping periods (21 versus 11) but this approach has the shortcoming of introducing double counting.

5.3.5 Measuring Overall Rating Mobility

The primal focus in evaluating or comparing credit rating transition matrices should be the 'mobility' as opposed to the 'stability' characteristics. The latter refers to the diagonal probability mass whereas mobility (or dynamics) refers to the off-diagonal probability mass. The reason is that one is more interested in the off-diagonal transition probabilities because mispredicting them implies greater economic costs. Thus an appropriate metric for transition matrix comparison should account for the diagonal dominance that is typical in sovereign credit migration matrices. In the literature, this comparison is usually based on L^1 (Israel et al., 2001) and L^2 (Bangia et al., 2000) Euclidean distances and eigenvalue/eigenvector analysis.[18] To test the validity of the homogeneity assumption, Arvanitis et al. (1999) compare different horizon matrices using a ratio of two matrix norms that evaluates the overall distance between their eigenvectors. Schorrocks (1978)

[18] For a survey of the metrics used for transition matrix comparison see Jafry and Schermann (2003).

and Geweke et al. (1986) proposed indices of mobility for Markov matrices based on eigenvalues and determinants. A pitfall of these metrics is that they typically cannot provide a clear signal regarding matrix comparison because they are inflated by the large concentration of probability mass along the diagonal (i.e. by the stability characteristics). To circumvent this problem Jafry and Schuermann's (2004) propose a mobility measure based on singular value decomposition, which we employ. This is outlined below.

Subtracting the identity matrix (I) from the transition matrix at hand gives the dynamic part of the system

$$\tilde{P} = P - I$$

or *overall mobility matrix* because I represents a fully stable (no migrations) system. The Jafry-Schuermann mobility estimator is defined as

$$m(\tilde{P}) \equiv \frac{1}{K}\sum_{i=1}^{K}\left\{\sqrt{e_i(\tilde{P}'\tilde{P})}\right\} \tag{5.21}$$

that is based on the singular value decomposition of the mobility matrix $\tilde{P}$ and where $e_i(\tilde{P}'\tilde{P})$ denotes the i^{th} eigenvalue of $\tilde{P}'\tilde{P}$.[19] Given the mobility matrices $\tilde{P}_A$ and $\tilde{P}_B$ where A and B are two estimators (for P), the differential measure

$$\Delta m(\tilde{P}_A, \tilde{P}_B) = m(\tilde{P}_A) - m(\tilde{P}_B) \tag{5.22}$$

has the nice property of largerly reflecting the concentration of the *off-diagonal* probability mass. Put differently, in the case where P_A and P_B have the same diagonal, the typical differential metrics are very likely to be zero whereas $\Delta m(\tilde{P}_A, \tilde{P}_B) \neq 0$. If the off-diagonal probability mass is diluted across a number of off-diagonal entries, then (5.21) is smaller than when it is concentrated in a few specific positions.[20]

[19]The use of the singular values of $\tilde{P}$ (i.e. eigenvalues of $\tilde{P}'\tilde{P}$) rather than its eigenvalues to ensure positivity of the metric, which is an essential property for the latter to stand as a distance-metric. Note that for the typical dimensions of credit migration matrices closed-form expressions for the singular values are intarctable. The latter is solved numerically using a Gauss code.

[20]The metric does not distinguish, however, between mass concentrated closer or farther from the diagonal which would be a desirable property as distant transitions imply greater mobility and higher economic costs.

5.3.6 Bootstrapping the Rating Migration Process

Constructing asymptotic confidence intervals for the estimated transition probabilities is difficult in a continuous framework. This is because there is no simple closed-form expression for the asymptotic variance-covariance of the intensity matrix estimator. Besides, the sovereign rating samples available are relatively small (few transitions) and so asymptotic estimators may not behave well in this context. Regarding the mobility differential estimator, equation (5.22), no asymptotic theory has been developed as yet. Given these shortcomings, we adopt Christensen et al.'s (2004) parametric bootstrap framework to compare alternative transition matrix estimators.

The DM estimator is compared with the HHR estimator by means of the following experiment. The continuous intensity matrix, Λ, and one-day transition matrix, $P = \exp(\frac{1}{365}\Lambda)$, are estimated from the observed rating histories. These estimates are used to construct artificial daily-rating data. For this purpose, the rating histories of the various obligors are conceptualized as independent realizations of a continuous, homogeneous Markov model.[21] We thus construct R bootstrap samples $\{B_j\}_{j=1}^{R}$ such that each B_j contains the same number of sovereign histories, N, as the observed sample and each sovereign's lifetime h and initial rating X_0 are as in the observed sample. We should stress that h varies across sovereigns (i.e. we have $h_i, i = 1, ..., N$) for two reasons: a) the rating agencies have included many emerging market sovereigns only recently, b) a rating history may terminate promptly in the case, for instance, of rating withdrawal. These issues are further discussed below in Section 5.4.

Each sovereign rating history is constructed as follows. Suppose X_0=Baa for the sovereign at hand and that the probability of transition from rating Baa to Aaa, Aa, A, Baa, Ba, B, C and D is $\hat{p}_{41}$, $\hat{p}_{42}$, $\hat{p}_{43}$, $\hat{p}_{44}$, $\hat{p}_{45}$, $\hat{p}_{46}$, $\hat{p}_{47}$ and $\hat{p}_{48}$, respectively, with $\left(\sum_{j=1,j\neq 4}^{8} \hat{p}_{4j} = 1\right)$. We transform $\hat{P}$ into cumulative probability ranges for the initial

[21] Cross-sectional homogeneity is essential because conditioning the transition matrix on the sovereign's characteristics would, in effect, imply smaller sample sizes and so less reliable inferences.

rating. The first range is $(0, \hat{p}_{41}]$. Summing $\hat{p}_{41}$ and $\hat{p}_{42}$ gives the cumulative probability that the new rating is either Aaa or Aa and hence, the next probability range is $(\hat{p}_{41}, \hat{p}_{41} + \hat{p}_{42}]$. The next range is $(\hat{p}_{41} + \hat{p}_{42}, \hat{p}_{41} + \hat{p}_{42} + \hat{p}_{43}]$ and so forth until eight ranges are computed, one for each possible transition. The next daily rating, X_1, is obtained by randomly drawing from a uniform distribution, $r \sim i.i.d.U[0,1]$, and then matching the draw with one of the above cumulative probability ranges. For instance, if $r \in (\hat{p}_{41} + \hat{p}_{42}, \hat{p}_{41} + \hat{p}_{42} + \hat{p}_{43}]$ then $X_1 = A$. Another uniform random draw gives X_2 and so forth until the end of this sovereign's lifetime so that $\{X_1, ..., X_h\}$ is obtained. This simulation is conducted independently for the remaining sovereigns to construct the bootstrap sample B_j that contains N rating histories. After iterating this simulation R times we have the bootstrap data sets $\{B_j\}_{j=1}^{R}$. Then each B_j is transformed into N rating transition histories and rating durations. On the basis of the latter, we estimate the one-year transition matrix $\left\{\hat{P}_j\right\}_{j=1}^{R}$ using the DM and HHR approaches, equations (5.5) and (5.16), respectively. For each matrix we calculate (5.21) and then the mobility differential estimator $\widehat{\Delta m}_j$. Thus we obtain an $R \times 1$ vector of default probability estimates per rating (last column entries in $\hat{P}_j, j = 1, ..., R$) or equivalently, the bootstrap distribution of the one-year probability of default estimator (PD hereafter). Similarly, we also have the empirical distribution of the mobility differential (Appendix 5.1 presents the bootstrap in steps).

The bootstrap distribution of the PD for each rating facilitates confidence intervals to compare the efficiency of the estimators and to test hypotheses about differences in the mean value. One can also assess the finite-sample bias by comparing the latter with the true (sample estimated) PDs used in the bootstrap DGP. The bootstrap PDs from each estimator are also compared by plotting their kernel density. Likewise, the bootstrap distribution of the mobility differential estimator (5.22) facilitates further comparisons across estimators. We use $R = 1000$ following Efron and Tibshirani (1993) who argue that this generally suffices to obtain reliable bootstrap confidence intervals. Nevertheless, in order to assess whether 1000 simulations guarantee convergence of the bootstrap, we experimented by performing two sets of 500 simulations. The estimated PD distributions using the two smaller (500) samples were virtually identical and comparable to those from the larger sample.

Similar bootstrap simulations are conducted to compare the homogeneous (HHR) and non-homogeneous (NHHR) hazard rate estimators. In order to introduce heterogeneity in the DGP for ratings, we estimate an intensity matrix per year on the basis of the observed rating transitions. Using the matrices $\hat{P}_t = \exp(\frac{1}{365}\hat{\Lambda}_t), t = 1, 2, ..., T$, a bootstrap daily-rating history is obtained for each sovereign as explained above. To make the comparison of the HHR and NHHR estimators more informative, equations (5.16) and (5.18), are deployed so as to evaluate the probability of transitions that occur over a 2-year horizon — in the former $\Delta t = 2$ and in the latter $t_1 - t_0, t_2 - t_1, etc.$ are 2-year horizons. We were reluctant to introduce heterogeneity in the simulations across shorter periods than one year for two reasons. First, the generator matrix estimator, $\hat{\Lambda}_t$, over 6- or 4-month periods would be highly inaccurate because few sovereign transitions are generally observed in such short intervals. Second, the anecdotal stability of the sovereign ratings suggests that there is within-year homogeneity in the underlying Markov process.

5.4 The Dataset

The data are from the *Moody's Default Risk Service* database that contains complete rating transition histories for all obligors (corporate and sovereign) since 1914.[22] Moody's provides ratings for each individual debt issue of about 33,463 issuers categorized by domicile and industry. The industry categorisation includes among others: banking, finance, industrial, insurance, public utility, real estate and sovereign. The assigned credit

[22]S&P's and Moody's ratings are highly correlated. For instance, out of the 49 sovereigns rated by both Moody's and S&P's in September 1995, 28 had equal rating whereas for those with different ratings, the gap was only one notch, with 7 exceptions that were 2 notches apart (Cantor and Packer, 1996).

rating represents Moody's assessment on the likelihood of each issuer to honour any type of future debt payment. Structural features of the debt issues such as maturity, coupon structure, collateralization and seniority are all taken into account in the assessment. We consider only sovereigns which had foreign-currency bonds outstanding and rated by Moody's some time during March 5, 1981 through March 4, 2004. Thus the rating histories of the 72 sovereigns in the sample have different lengths.[23] For the DM approach, based on year-on-year ratings, the sovereigns are observed on the 5^{th} March.

Our empirical analysis is based on the *issuer's* foreign-currency rating history rather than on the history of every single foreign-currency bond issue. Moody's occasionally assigns different ratings to the bonds of the same issuer depending on their characteristics.[24] In order to convert the sovereign bonds' ratings into a single rating for the sovereign at any point in time we observe its lowest rating on senior unsecured bonds which have not yet matured or been repaid. The latter is the most meaningful indicator of a sovereign's likelihood of defaulting on any of its bonds (Moody's, 2003). For each sovereign, the exact rating transition dates are recorded.

A sovereign default is defined to occur whenever, according to Moody's records, a country defaults on any of its foreign-currency rated bonds. Moody's does not have a 'default' rating category as such but it records default dates. While in default an issuer is still rated (e.g. as Caa) and the rating representing the severity of default. For each country, we treat the date of the first default announcement as the single default date — these

[23]We exclude countries that have been assigned ratings for domestic currency bond issues as well as for types of foreign currency debt issues other than bonds (i.e. foreign currency bank deposits, country ceilings). Excluding also sub-sovereigns and municipals (local and regional) as well as bonds with convertible features, the 72 sovereigns have had around 3,000 registered debt issues with the aforementioned characteristics over the period. Including all rating changes for every issue we have a total of 6,056 rating transitions experienced by 68 of the 72 sovereign issuers. The transitions per country range from just 1 for some small sovereigns (Bermuda, Estonia, Latvia, Mauritius, Morocco, Oman) up to 961 for Sweden. Sovereigns that have never experienced a transition on any of their foreign bond issues are Chile, Guatemala, Egypt and El Salvador, which were first rated in 1999, 1997, 2001 and 2002, respectively.

[24]For instance, Russia's 'MinFin' US dollar bonds have been generally rated lower than its Eurodollar bonds. During the 1999 Russian crisis, defaults occurred on the former but not on the latter.

are 17/08/98 (Russia), 30/11/98 (Pakistan), 25/08/99 (Ecuador) , 20/01/00 (Ukraine), 7/09/00 (Peru), 13/06/01 (Moldova), 30/11/01 (Argentina) and 15/05/03 (Uruguay). The rating from which the transition to default took place and the ratings that were assigned as characterisation of the default status need to be disentangled. In order to identify the former we track the rating sequence up to the default date and throw away the rating transitions that occurred very close (within a month) to it. For instance, Ukraine was downgraded from Caa3 to Ca at day -15 in default event time so the pre-default rating is Caa3 as the downgrade clearly reflected the pending default which was only delayed.

Appendix 5.2 shows the number of foreign-currency sovereign bond issuers rated by Moody's from 1981-2004 and provides a breakdown according to issuer's characteristics, geographical location, state of the economy and credit quality. Most industrialised countries have ratings from the beginning of the sample period (they dominate the sample until the mid 1990s) whereas many emerging economies were first rated later within the sample window. More specifically, only 11 sovereigns were rated over 1981-1986 which comprised 9 industrial and 2 emerging economies whereas 17 industrial and 55 emerging markets were rated in 2003. Regarding non-investment grade ratings, none was observed before 1987, between 2 and 18 during 1987-1997 and around 25 during 1998-2004.

As aforementioned, the initial date of the rating history varies across sovereigns. The same applies to the termination date or 'right' censoring in the terminology of survival analysis. Termination can occur either at the end of the time window (i.e. March 4, 2004) called implicit censoring or at an earlier point due to withdrawn rating or default called explicit censoring. A sovereign receives an withdrawn rating (WR) status if the underlying debt issues are no longer outstanding, i.e. bonds have matured, been repaid or called. Most often the latter reflects the issuer's temporary exit from the public bond market rather than having negative credit implications (Carty, 1997). However, sovereign WRs are rather scarce because sovereigns, as opposed to corporates, rarely retire all their debt simultaneously — our sample contains only 5 countries that have, at any time, experienced a WR.[25] We follow the literature and consider WRs as non-informative, and so they are excluded from the rating history of the specific sovereign. The sovereign is treated as a new independent issuer when the rating is later resumed. Likewise, any rating assigned between the default date and the end of the default episode is discarded. The latter is defined as the date when the sovereign exceeds the B3 rating. The sovereign is treated as a new independent issuer when the default episode ends.[26] Half of the 8 default countries (Pakistan, Peru, Russia and Ukraine) recover from default and re-emerge in our sample as new issuers whereas the rest (Argentina, Ecuador, Moldova and Uruguay) remain in default at the end of the sample window.

Finally, in order to focus on the broader ratings Aaa, Aa, A, Baa, Ba, B, C (and default) used by Moody's prior to 1982, we label the numbered sub-categories according to the mother category, e.g. Baa1, Baa2, Baa3 are treated as Baa. The lowest sub-categories, Caa1, Caa2, Caa3, Ca1, Ca2, Ca3, Ca, contain either very few or no observations at all and so are all merged into a C category. There are two reasons for restricting our analysis to the eight coarse ratings. First, this reduces the number of parameters to be estimated and increases the sample size of transitions per rating. Second, the reliability of finer transition matrix estimates in credit risk applications is doubtful and thus the coarse rating system has emerged as the industry standard.

The above considerations result in an effective 1981-2004 sample of 81 sovereigns of which 4 and 5 countries have re-emerged from default and WR, respectively, and therefore are treated as new independent issuers. Figure 5.1 displays the aggregate ratings distribution over all issuers and years in the sample. Investment grade sovereigns represent 70% of the sample on average. Appendix 5.3 provides detailed information on the distribution

[25] Debt withdrawals are common for corporates, for instance, in the case of mergers or liquidation.

[26] This procedure is equivalent to treating the default state as absorbing.

of credit ratings in each year along with some summary statistics. In line with the above discussion the numbers exclude WRs from the date of withdrawal to the new rating date and leave out defaulted countries from immediately after the default date until recovery. As a whole, we have a total of 759 sovereign-year rating observations and 104 rating transitions.[27] A discrete estimation framework (ignoring within year rating activity) captures only 80 of these 104 transitions which means throwing out effectively 23% of the observed migrations.[28], [29]

5.5 Properties of Migration Matrix Estimates

In this section we compare the properties of rival transition matrix estimators. The statistical framework for the comparison is a parametric bootstrap. We first compare the 1-year probability matrices estimated from discrete and continuous type approaches. Second, we investigate the degree of discrepancy between the two estimators as the transition horizon increases from 1 to 3 years. Third, the added value of time-heterogeneous estimators when there is time variation in the underlying Markov process is also assessed. For the latter purpose, we introduce yearly heterogeneity and focus the comparison on 2-year transitions. The discussion focuses primarily on the relative bias and efficiency of the PD estimates and on the overall mobility or dynamics implied by the transition matrix estimates. Finally, the presence of duration dependence and rating momentum in rating transitions is investigated.

5.5.1 Discrete versus Continuous Estimators

We start by illustrating how the differences between the discrete and continuous estimators materialize in the Moody's dataset. In the DM approach, equation (5.5), entry p_{ij} of the

[27]The total number of annual observations is the sum of all yearly ratings per sovereign.

[28]The only available study on sovereign transition matrix estimation, Hu et al. (2002), uses only 26 S&P annual rating transitions during 1981-1998 in a discrete-time framework.

[29]The empirical analysis in this chapter is conducted using Gauss 3.4 and LIMDEP 8.

1-year transition matrix is estimated by adding the number of sovereigns rated i at the start of any year ($t = 1, ...T$) and rated j at the start of year $t+1$ and dividing the sum by the total number of sovereigns that started any year in rating i. In the continuous approach the entry λ_{ij} of the generator is provided by the HHR estimator, equation (5.15), which divides the number of transitions from i to j observed during the T-year sample window over the total time spent (in years) at rating i by all sovereigns. The 1-year transition probability matrix is then obtained as the matrix exponential of the generator, (horizon $\Delta t = 1$) for a 6^{th} order Taylor expansion. Table 5.1 reports the transition matrix estimates. Our DM estimate is very similar to that reported by Moody's (2003). Both transition matrices are diagonally dominant indicating rating stability. There is also heavy concentration around the diagonal, that is, most observed transitions are towards a neigbouring rating. Higher migration activity and more distant transitions are associated with migration from lower credit qualities (Ba, B and C categories). Another common characteristic is that for each row (initial rating i) the transition probabilities decrease as one moves farther from the diagonal. This is referred to as row monotonicity and is a general feature of credit rating migration matrices (J.P.Morgan, 1997, p.73). A violation of monotonicity occurs for the B rating such that there is a higher probability of migrating to default than to C. A possible explanation could be the noisy nature of the data for the low B rating. The above effect is more prominent in the multinomial estimator which is unsurprising because it ignores the intra-year rating activity.

The above discrete and continuous (homogeneous) estimators differ in three respects. First, the HHR probability estimates are positive for most transitions even if they have not been observed. For such cases, the DM estimates are zero. The HHR method spreads the off-diagonal probability over almost all ratings whereas the DM concentrates the probability mass around the diagonal. Second, the HHR transition matrix estimate exhibits greater migration volatility for low ratings, that is, relatively less diagonal probability mass. One reason for this difference is that continuous estimators better capture the rating dynamics. Third, the PD of a C-rated sovereign suggested by the HHR method is relatively high. A possible rationale is that the DM method will only record a default from C when the sovereign starts the year at C and ends it in default. However, the duration of C is short, in most cases less than a year, as it is just a transitional rating toward default. The transition from C to default will not be captured by the DM, which will consider it as a default from a previous rating (e.g. from B). On the other hand, it will be recorded by the continuous estimator. The short durations in C peak the default intensity from C leading to a higher estimate from the continuous method.

We now deploy the parametric bootstrap technique described in Section 5.3.6 to compare the finite sample properties of the DM and HHR estimators. For this purpose we use the homogeneous Markov model characterized by the generator matrix and the associated transition matrix estimates (Table 5.1). The latter are referred to as true default probabilities. To preserve space and without loss of generality we focus the discussion on the default probabilities.[30] Figure 5.2 plots the kernel (Gaussian) density of the bootstrap 1-year PD estimates for the investment grade categories. The bold vertical line signifies the true PD and the two dashed lines indicate the 95% confidence interval. For the top ratings, Aaa, Aa, A, the kernel density for the DM estimates is not plotted because these are zero for all simulated paths. Table 5.2 shows, for each rating, summary statistics for the bootstrap PD estimates along with the true PD value. For the three highest ratings (Aaa, Aa, A) with true default probabilities as small as 1.89×10^{-6}bp, 0.000158bp and 0.0556bp, the DM estimates are zero for all simulated paths. The distribution of the HHR estimator is roughly exponential with upper 97.5% quantiles of 1.23×10^{-5}bp, 0.0009bp and 0.238bp, respectively and means quite close to the 'true' default rate (small bias). Transitions to default from high ratings are not observed in the sample or in the simulated paths. However, the continuous method is able to provide an estimate of how rare such

[30]Simulation results for all transition probabilities are available from the authors upon request.

events are.

Regarding the Baa rating, the DM estimator produces a zero PD for most paths and a PD in the range [60, 125]bp for a few paths (Figure 5.2) which leads to a rather high upper quantile of 75.2bp. In sharp contrast, the upper quantile of the HHR estimator's distribution is 9.72bp which means a notable increase in accuracy compared to the former estimator. The higher DM quantile is because about 2.5% of the simulated paths contain at least one transition from Baa to default, $0.0223\% \times 112 = 0.025$ where 112 is the number of observed Baa ratings at year-beginning (N_{Baa}) as shown in Appendix 5.3. If one default occurs, then the DM probability estimate in simulation j is $1/N^j_{Baa}$ where $E(N^j_{Baa}) = 112$ and $1/112$ is roughly 75.2. Our analysis can be used to assess the adequacy of the minimum probability at 3bp that has been established by the Basel Committee for unobserved (thus far) events. This threshold probability clearly falls in the HHR confidence interval for Baa-rated sovereigns while it is well beyond the 97.5% quantile of the HHR distribution for the Aaa, Aa and A ratings. Hence, the Basel II threshold appears too 'tough' for these higher ratings.

The counterpart results for the lower ratings (Ba, B, C) are set out in Figure 5.3. We observe that the DM probability estimator has a smooth distribution only for the B category. This is not surprising given that direct default migrations from B are expected to be the most frequent. The total number of observations over the period are 132 for Ba, 85 for B and 6 for C so according to the true default probabilities the expected number of default migrations per simulation are $83.07\text{bp}\times 132 = 0.010$, $487.2\text{bp}\times 85 = 0.041$ and $2897\text{bp}\times 6 = 0.017$, respectively. The DM estimator is similar to the HHR estimator in terms of accuracy for the B rating whereas the latter is more efficient (tighter confidence intervals) for the Ba and C categories. The trimodal distribution of the HHR estimator for the Ba category is attributed to the fact that there are two main transitions from Ba in the true generator matrix, one is an upgrade from Ba to Baa , with 6.76% probability, and the other is a downgrade from Ba to C, with 8.55% probability. This creates a division of the simulations between those where the upgrade occurs and those where the downgarde occurs. When the up(down)grade occurs the PD de(in)creases significantly. In the case of no transition we obtain the middle peak, that is the PD directly from Ba which is 0.08% in the true transition matrix.

In terms of bias, measured by the difference between the true PD and the mean of the bootstrap distribution, the HHR estimator shows less bias than the DM estimator for the higher ratings whereas the opposite holds for the lower ratings. Moreover, as noted earlier, the HHR estimator captures the frequent transitions with short duration of the low categories (on the way down to default) resulting in comparatively higher PD estimates. This is in line with the finding that for the B and C ratings the DM and HHR estimators exhibit negative and positive bias, respectively.[31]

We turn now to the issue of whether the difference between the DM and HHR default probabilities is statistically significant. For this purpose, we conduct two-sided bootstrap tests for H_0:$PD_{HHR} - PD_{DM} = 0$. Table 5.3 (panel A) reports the summary statistics for the probability differential measure, $(\hat{PD}_{HHR} - \hat{PD}_{DM})$, over replications. It turns out that all the 95% confidence intervals contain zero with the exception of the rating A for which the above null is rejected — for this mid rating, the DM default probability estimate is zero because of the zero observed default transitions from A in the sample whereas the more efficient HHR estimator gives a non-zero probability because it jointly exploits the information that there are transitions from A to Ba and Baa and from Ba to D. Therefore we infer that the HHR estimator can stand out as less biased than the DM estimator for some mid ratings.

Finally we address the question of whether there are significant differences in the over-

[31]For the default probability from the Aaa, Aa and A ratings, the negative bias in the DM estimator is explained by the fact that it predicts 0 whenever no transitions are observed in the bootstrap sample, i.e. $N_{ij}(t, t+1) = 0$ for $i = Aaa, Aa, A$ and $j = D$.

all ratings mobility (or migration risk) implied by the DM and HHR methods. Table 5.3 (Panel B) provides the sample mobility differential, $\widehat{\Delta m}(\tilde{P}_{DM}, \tilde{P}_{HHR})$, as well as summary statistics of the bootstrap distribution. First, the DM matrix implies higher overall mobility than the HHR matrix and the difference is statistically significant as suggested by the 95% confidence interval not containing zero. This suggests that the concentration of probability mass at off-diagonal positions in the DM transition matrix is higher than that in the HHR counterpart. Upon closer inspection (Table 5.1) it is apparent that the DM matrix is relatively sparse (large number of zero entries) and its transition probability mass is largely concentrated in those few ratings for which transitions have been observed in the sample (i.e. around the diagonal). Default probability mass is higher, in general, for the HHR. However, the HHR matrix spreads the PD mass over more off-diagonal positions or equivalently, a larger number of ratings. These results are consistent with the discussion in Section 5.3.5, namely, that the presence of a few large off-diagonal terms inflates the m metric considerably.

5.5.2 Increasing the Transition Horizon

The DM and HHR transition matrix estimators are expected to show more marked differences when the horizon is beyond one year because the latter will allow for more rating activity. Thus comparing the three estimators over longer horizons bigger differences are likely to bear out. The probability matrix estimates for the 2- and 3-year transition horizons are presented in Table 5.4. It turns out that the DM estimator captures default risk only for one or two ratings, Baa and B, in contrast with the HHR estimator. One can notice that the DM estimate of the PD for Baa is about twenty-fold above the HHR estimate — the 2-year default probabilities are 192bp and 9bp and the 3-year are 333bp, 22bp for the DM and HHR cases, respectively. The smaller HHR default probability for Baa is more plausible because sovereigns spend relatively long times in the mid Baa state (versus other ratings) on their way up(down) the rating scale — the latter is reflected in the denominator of equation (5.15) — which will thus pull down the default probability for Baa. The differences are striking at the lowest end of the rating spectrum — for rating C the DM default probability estimate is zero throughout whereas the HHR estimates are 38.18bp and 35.49bp for the 2- and 3-year horizons, respectively. The zero DM estimate is unrealistic since there are reasons to believe that there is a non-zero probability of migrating from C to default over a relatively large horizon of 2 or 3 years given a large enough sample of sovereigns. This findings are just a reflection of a shortcoming of discrete versus continuous transition probability estimators, namely, the relatively short stay of sovereigns in the low rating C biases the former but not the latter because it accounts for duration through the intensity matrix. The upshot is that the DM method underestimates the default probability for the highly risky sovereigns.[32]

Regarding the overall mobility metric, m, it turns out that the DM estimator gives higher values than the HHR estimator for all horizons ranging from one to four years. Unsurprisingly, the differential $\widehat{\Delta m}(\tilde{P}_{DM}, \tilde{P}_{HHR})$ increases with the time horizon.

5.5.3 The Time-homogeneity Assumption

We now assess the validity of the homogeneity assumption for the sovereign rating process by comparing the performance of the HHR and the NHHR approaches. It seems plausible that the longer the horizon, the more apparent the heterogeneity is and so we focus the comparison primarily on the 2- and 3-year horizons. Table 5.5 presents the 2- and 3-year horizon probability matrices from the NHHR approach. At first glance, the difference between the above two hazard rate estimators is not as striking as that between the two homogeneous, one discrete (DM) and the other continuous (HHR), estimators. The NHHR estimator yields non-zero transition probability estimates in positions that correspond to rare transitions (sample) but to a lesser extent than the HHR estimator.

[32]The DT estimates $\hat{p}_{C,B} = 1$ and $\hat{p}_{C,Ba} = 1$ (for the 2- and 3-year horizons, respectively) stem from one transition — Romania migrates from C to B after 2 years and to Ba after 3 years. All other C-rated issuers have durations shorter than 2 years and so their transitions are not captured by the DM method.

Low credit ratings exhibit greater overall migration in the HHR matrix whereas the opposite holds for the NHHR matrix. Furthermore, there are discrepancies between the two estimators in the default probability particularly so for the Baa and the lowest category C, with the NHHR estimator giving lower probabilities.

It turns out that increasing the transition interval from 2 to 3 years exacerbates the differences between the HHR and NHHR matrix estimates. This is corroborated by the $\widehat{\Delta m}$ statistic which increases with the transition horizon. The overall mobility measure $\widehat{m}$ is larger for the HHR method and this may be because this approach is applied directly over the whole sample whereas in contrast the NHHR transition estimate, in order to control for heterogeneity, is an average of the sequential estimates over each (2 or 3 year) interval. Thus the former estimator exploits more of the (indirect) transitions between distant ratings in the sample and thus has higher off-diagonal probability mass.

We now statistically compare the two continuous estimators on the basis of the bootstrap replicates. To preserve space we focus on default probability estimates over a 2-year horizon that we think is sufficient for time variations in the rating process to become apparent. A different generator is used for each year (to introduce heterogeneity), as discussed in Section 5.3.6 and the true PD over the sample window is the issuer-based weighted average of all the 2-year transition matrices. The kernel density and summary statistics of the bootstrap default probabilities for investment grade issuers can be seen in Figure 5.4 and Table 5.6. For the Aaa and Aa ratings with zero true PD, the kernel density of the NHHR estimates is not plotted because these are zero for all simulated paths. The latter can be rationalized as follows. Direct transitions to default do not occur from high ratings. The two continuous estimators facilitate default intensities for the latter cases by capturing indirect transitions. The difference between them stems from the homogeneous versus heterogeneous aspect. The HHR transition matrix, $\hat{P}$, is based on continuous transition intensities obtained directly from the whole rating histories. In contrast, the NHHR counterpart matrix is the average of matrices, $\hat{P}(t_j, t_{j+1})$, estimated sequentially over non-overlapping 2-year histories. Hence, the NHHR estimator will naturally combine fewer indirect transitions between distant ratings than the HHR estimator. To illustrate, suppose that sovereign X migrates from Aa to Baa, sovereign Y from Baa to B and sovereign Z from B to default in a 6-year period. The HHR estimator will combine the information and give a default probability from Aa to default whereas the NHHR that conditions on 2-year periods will not be able to combine all this information, unless it happens within the 2-year period. This explains why the NHHR approach generally produces zero default intensities for the top ratings.

For the A and Baa ratings, with non-zero true PD, both estimators produce smooth distributions (Figure 5.4). In the case of the A rating, the NHHR estimator yields zero probabilities for most simulated paths, and very few paths with probabilities in the range $(0, 2]$bp, giving a 97.5% confidence band of 0.47bp. The corresponding upper quantile for the HHR estimator is 1.93bp. The NHHR estimator is more efficient and less biased (Table 5.6). For the Baa category, the 97.5% quantile of the bootstrap distribution for the NHHR estimator is about sixteen-fold smaller that that for the HHR estimator. The NHHR estimator is again more efficient and less biased.

Figure 5.5 plots the kernel density of the PD estimates for non-investment grade issuers. All of them are centered at the true default rate (c.f. true and mean PD in Table 5.6), except that of rating C. However, the NHHR estimator produces tighter confidence intervals than the HHR estimator (more efficient) for all ratings, the only exception being Ba for which the two are comparable. For the C rating the difference is remarkable. The HHR produces a very wide 95% band of [0, 8396bp] whereas the corresponding band for the NHHR is [411bp, 3522bp]. Moreover, the NHHR estimator has smaller bias for all ratings.

Table 5.7 reports the summary statistics for the bootstrap distribution of the 2-year

PD differential ($\hat{PD}^{HHR} - \hat{PD}^{NHHR}$) for all ratings. All the confidence intervals contain zero and thus, there is no evidence that the HHR and NHHR estimators yield significantly different PDs. However, in terms of overall mobility the two transition matrix estimates appear significantly different. More specifically, the mobility in the HHR matrix is significantly larger than that in the NHHR matrix which means more probability mass at off-diagonal positions. Put differently, disregarding heterogeneities might result in transition matrix estimates that suggest more frequent rating migration (possibly between farther apart ratings) than that actually implied by the underlying Markov process.

5.5.4 Robustness Checks

One may argue that the above comparison results may be determined by the sample period used to estimate the parameters of the data generating process for the simulations. We check the robustness of our findings by carrying out the same experiment using only the last six years in the sample (March, 1998- March, 2004). This period is characterised by the introduction of lower credit-quality (emerging market) sovereigns and the occurrence of all eight defaults. The post-1998 era is thus representative of substantially more frequent rating transitions.[33] As expected, the transition probability estimates for non-investment grade issuers rise dramatically relative to those for the overall sample period. However, the comparison results do not change qualitatively.

All three estimators under study assume that the group of individuals examined is homogeneous, i.e. $p^s_{ij}(t) = p_{ij}(t)$ for all sovereigns $s = 1, 2, ..., N$. But the validity of this assumption can be questioned for at least two reasons.[34] First, for some highly rated sovereigns the initial state perpetuates. For example, Austria has maintained the initial Aaa rating since 1977, France since 1992. Second, regional considerations or the level of

[33] Results for the restricted period are available from the authors upon request.

[34] The assumption of cross-section independence is also questionable because rating transitions could be correlated due to regional contagion, debt links, capital flows and IMF fund flows. S&P's (2004) argue, however, that their rating transitions over the recent past are country-specific.

developement of a country may increase the propensity of up(down)grade. For instance, Italy and Israel both had an A rating in 1995 but Italy is more likely to be upgraded than Israel. To control for the first, one can utilize a 'mover-stayer' type model that implies two distinct continuous Markov processes, one for movers and another for stayers. But maximisation of this model's likelihood function is rather complicated.[35] To gauge the robustness of our results in this regard, we measure the percentage of countries rated over a long period (beyond 15 years) and whose credit status has remained constant over the entire time window. Only 5 sovereigns meet these requirements in the sample — Austria (Aaa), France (Aaa), UK (Aaa), Australia (Aa) and Belgium (Aa). The transition matrix in this simple time-homogeneous, mover-stayer model is $Q = S + (I - S)P$, where Δt is the transition horizon, $S \equiv diag(s_1, s_2, ..., s_8)$ and $s_i, i = 1, ..., 8$, denotes the proportion of stayers among those countries whose initial rating was i and $P(\Delta t)$ is the HHR transition matrix. In our sample, $Q(\Delta t)$ and $P(\Delta t)$ only differ in the first (Aaa) and second (Aa) rows. The results suggest that the diagonal probabilities rise slightly whereas the off-diagonal probabilities in the first and second rows, decrease by a factor of s_1 and s_2, respectively, for all three estimators. Hence, the findings regarding the relative properties of the estimators are qualitatively similar.[36]

5.6 Detecting Non-Markovian Behaviour

The DM, HHR and NHHR estimators build on the premise of a Markov migration process. To the best of our knowledge, the plausibility of the latter has not been assessed in the context of sovereign debt. In this section, we attempt to fill this gap in two ways. First,

[35]It requires the use of the Expectation Maximization (EM) algorithm of Dempster et al. (1977). A mover-stayer model is applied to corporate bonds by Frydman and Kadam (2002). They find that for newly issued low-quality bonds (C rating), default probabilities are substantially smaller than what is implied by the usual Markov model and those issuers are more likely to stay in their current low rating.

[36]Detailed bootstrap results over the 1998-2004 period and the mover/stayer probabilities from each of the three estimators under study are available upon request.

we test whether the transition matrix estimate conforms to the homogeneous Markov structure by means of spectral analysis. Second, we investigate the presence of momentum and duration effects by means of panel logit models.

5.6.1 Testing for Markov Structure of the Transition Matrix

Let the 1×8 state vector $\mathbf{x}_t$ contain the probability of each credit rating for any sovereign at time t. Assuming homogeneous Markov evolution according to transition matrix P and an initial state vector $\mathbf{x}_0$ (e.g., $\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ indicates that the sovereign starts at rating Aa) then the state vector n periods ahead is given by

$$\mathbf{x}_n = \mathbf{x}_0 P^n \tag{5.23}$$

which can be rewritten as

$$\mathbf{x}_n = \mathbf{x}_0 \mathbf{E} Q^n \mathbf{E}^{-1} \tag{5.24}$$

by using the eigenvalue decomposition $P = \mathbf{E} Q \mathbf{E}^{-1}$ where Q is the diagonal matrix of eigenvalues and $\mathbf{E}$ is the eigenvector matrix.

Since each row of P sums to one, it follows that there exists a trivial unity eigenvalue which is the largest in magnitude ($q_1 = 1$). This has implications for the rating process. Specifically the process is neutrally stable, namely, when the time elapsed tends to infinity it reaches the steady state

$$\mathbf{x}_\infty \equiv \lim_{n \to \infty} (\mathbf{x}_0 P^n) = \mathbf{x}_0 \mathbf{E} Q^\infty \mathbf{E}^{-1} \tag{5.25}$$

and as time goes to infinity, all eigenvalues but q_1 go to zero. Hence, the steady state vector $\mathbf{x}_\infty$ is a multiple of the eigenvector associated to q_1.[37] The rate at which the process decays to the steady state is dictated by the second largest eigenvalue (q_2) called the transient decay term — the smaller q_2 is the faster the process reaches the steady

[37] The results regarding the unity eigenvalue and the corresponding eigenvector of Markov transition matrices can be proven using the Perron-Frobenius theorem for matrices.

state. It can also be shown that the eigenvector associated to q_2 provides the asymptotic distribution of survivors (sovereigns not ending in default) and thus it provides insights on the rating towards which the survivors will converge asymptotically.[38]

It follows from equations (5.23)-(5.25) that, for a process to be homogeneous Markov, the eigenvalues of P associated with increasing horizons should decay exponentially and that the eigenvectors should remain constant. In other words, if q_{2t} is second largest eigenvalue of the migration matrix for transition horizon t, then $ln(q_{2t}) = -Ct, C > 0$. For a Markov rating process, this log-linear relationship has to hold for all subsequent eigenvalues. The second to fifth largest eigenvalues of $\hat{P}$ for varying transition horizons from 1-4 years are plotted in Figure 5.6. The estimation of the transition matrix is based on the HHR method. The graph strongly supports the log-linear relationship above and consequently the Markov properties of the rating process.

Figure 5.7 shows that the 2nd largest eigenvector of $\hat{P}$ is very similar across horizons which provides additional evidence for Markovian behaviour. Moreover, the rating distribution exhibits a peak at Aaa. This suggests that the survivors in the long-term tend to settle at the highest rating. We carried out robustness checks for this analysis by constraining the sample to the last 6 years (1998-2004) and the results did not change qualitatively.

5.6.2 Testing the Duration Effects Hypothesis

To test the effect of duration on credit rating transitions we estimate panel logit models using monthly sovereign credit ratings over 23 years (March, 1981-March, 2004).[39] The

[38]The steady state solution is equal to the absorbing row of P, that is, for any transition matrix exhibiting at least one non-zero probability of default the state vector will settle at the default state. Put differently, given enough time and a constant (homogeneous) migration matrix all sovereigns will end up in default. Since the rate of decay to the steady state of the sovereign migration process is very slow (long durations) the time homogeneity assumption is unlikely to hold over such long time periods. Hence, the economy and consequently the migration process often changes long before the default state is reached.

[39]The duration effect is different from the 'ageing' effect examined by Altman and Kao (1992) for corporates. The latter refers to the relation between the time since issuance of individual bonds and the

duration measure, d_{it}, is obtained for each sovereign at the end of each month as the time elapsed since the last transition to the current state. To illustrate, consider a sovereign that experienced a rating transition to Ba in June 2002, then to B in September 2003 and has not moved since then. The duration in, say, June 2003 is 12 months and in March 2004 it is 6 months. We assume that the rating histories start at the beginning of our sample window in March 1981 (left-censored durations) which is not too restrictive because very few issuers had been rated before 1981. The effect of duration on the rating transition probability is assessed separately for upgrades (UP) and downgrades (DW). To do so, we define the endogenous variables

$$UP_{it} = \begin{cases} 1 & \text{if sovereign } i \text{ was upgraded in month } t \\ 0 & \text{otherwise} \end{cases}$$

$$DW_{it} = \begin{cases} 1 & \text{if sovereign } i \text{ was downgraded in month } t \\ 0 & \text{otherwise} \end{cases}$$

and estimate the following logit regression for each

$$y_{it}^{*} = \alpha + \beta d_{it} + \gamma' \mathbf{z}_t + \varepsilon_{it},\ \varepsilon_{it} \sim iid(0, \sigma_i^2),\ \ i = 1, ..., N,\ \ t = 1, ..., T \tag{5.26}$$

where y_{it}^{*} is a continuous latent variable such that $UP_{it} = 1$ for $y_{it}^{*} \geq 0$ and $UP_{it} = 0$ otherwise (likewise for DW_{it}) and $\mathbf{z}_t$ is a 7×1 year dummy vector over 1998-2004 that controls for the fact that many emerging economies entered the sample after 1997. If durations are more stable over time for industrial countries, then the logit error variance should be smaller for this group. To control for this groupwise heteroskedasticity, the error variance is allowed to differ between industrial and non-industrial countries according to $\sigma_i = [\exp(\psi + \xi r_{it})]$ where $\exp(\psi) = \frac{\pi}{\sqrt{3}}$ and $r_{it} = 1$ if i is industrial, ψ and ξ are time-invariant parameters.

Appendix 5.4 summarises the observed rating durations. Two features emerge. First, average duration increases with rating quality, the only exception being the absorbing

upgrade (downgrade) probability.

default rating. This can be explained in terms of the low transition probabilities for high credit-quality sovereigns — using survival theory, it can be shown that the expected duration in state i is negatively related to the probability of transition away from i. Second, the standard deviation of duration is also larger for the high credit-quality ratings but this is just a reflection of their relatively large durations.

We now address the question of whether duration influences the upgrade/downgrade probability. Table 5.8 reports the logit estimation (ML) results.[40] The duration coefficient is negative for both upgrades ($\hat{\beta} = -1.88$) and downgrades ($\hat{\beta} = -1.66$) which suggests that transition probabilities are negatively influenced by duration — the more time a sovereign spends in the current rating the less likely it is to be migrated. These findings are consistent with those in Lando and Skodeberg (2002) for corporates and can be explained in terms of the common practice by rating agencies of upgrading/downgrading gradually notch by notch. The latter results in short durations and high migration risk for the low end of the rating scale and vice versa. Finally, note that the estimated error variance is higher for industrial countries ($\hat{\xi} > 0$) albeit not significantly. The positive parameter is an artefact of the high durations that inflate the variance of higher ratings. The standardised durations are comparable across ratings, tehreby justifying the insignificant result (see Appendix 5.4).

5.6.3 Testing the Rating Momentum Hypothesis

In order to test for momentum effects in sovereign rating migrations we define

$$UM_{it} = \begin{cases} 1 & \text{if sovereign } i \text{ was upgraded to the current rating over } [t-1, t-23] \\ 0 & \text{otherwise} \end{cases}$$

[40] Our primary focus is on testing effects that are 'internal' to the rating process and would automatically violate the Markov assumption. Nevertheless, equation (5.26) was reestimated by substituting a business cycle dummy for d_{it}. We find that the downgrade probability is significantly higher in recessions, however, the result is marginal.

$$DM_{it} = \begin{cases} 1 & \text{if sovereign } i \text{ was downgraded to the current rating over } [t-1, t-23] \\ 0 & \text{otherwise} \end{cases}$$

which are referred to as the upward and downward momentum indicators, respectively (and t denotes months).[41] First, we estimate the upgrade logit regression

$$y_{it}^{*} = \alpha + \beta v_{it} + \gamma' \mathbf{z}_t + \varepsilon_{it}, \; \varepsilon_{it} \sim iid(0, \sigma_i^2), \quad i = 1, ..., N, \quad t = 1, ..., T \tag{5.27}$$

where $UP_{it} = 1$ for $y_{it}^{*} \geq 0$ (as defined in the previous section) and $v_{it} \equiv UM_{it}$. A similar logit is estimated to test for downgrade momentum. The exogenous vector $\mathbf{z}_t$ and variance σ_i^2 are as defined in the previous section.

The estimation results are presented in Table 5.9. The downgrade logit estimates provide strong support for the momentum hypothesis with a highly significant coefficient at $\hat{\beta}$=1.20. This suggests that a downgrade in the previous two years significantly increases the current downgrade probability. These findings are in line with the extant evidence for corporate debt (Lando and Skodeberg, 2002; Nickell et al. 2000). Moreover, the residual variance is significantly lower for industrialized countries ($\hat{\xi} = -0.17$) which is in line with the more stable downgrade momentum history for these economies. In sharp contrast, the upgrade logit provides no evidence of momentum.

5.7 Conclusion

Sovereign credit ratings and the associated migration probabilities play a major role in modern credit risk management, valuation and international capital allocation. Different models and estimators for rating migration matrices have been proposed. However, the extant methodologies have been mainly applied to corporate credit risk. Very little is known on the finite-sample properties of these estimators in the context of sovereign

[41] Let R_{it} denote the rating for country i at period t. To construct them we initially set $UM_{it} = 0$ and compare the $R_{i,t-1}$ with $R_{i,t-j}$ for $j = 2, ..., 23$ sequentially. For instance, if $R_{i,t-j} < R_{i,t-1}$ for $j = 2$ then an upgrade occurred at $t-1$ and $UM_{it} = 1$ and the comparison stops, otherwise $j = 3$ and so forth.

ratings. There is substantial evidence that sovereign and corporate ratings behave differently. For instance, the former are notably more stable than the latter and so it is not clear that continuous and possibly time-heterogeneous estimators will perform better than the conventional DM estimators in the sovereign context.

This chapter contributes to the literature by comparing the finite-sample properties of three different credit migration estimators — the DM approach considered as the industry standard and two continuous, hazard rate approaches that differ in how they treat time-heterogeneity. The sample for the analysis is an unbalanced panel of ratings for 72 sovereigns over the 1981-2004 period. The comparison is conducted through a parametric bootstrap method that facilitates empirical confidence intervals and bias measures. The three estimators have in common that they build on the Markov property for the rating evolution which implies that the future rating is independent of the rating history. In a panel logit framework, tests are conducted for the presence of two non-Markov effects in the sovereign rating process known as rating momentum and rating duration. Such effects induce time-heterogeneity in the rating process and thus, a hazard rate estimator that allows for the latter merits particular attention.

Significant differences are found between discrete and continuous type estimators, on the one hand, and between homogeneous and heterogeneous estimators, on the other. The continuous estimators yield more accurate default probabilities than the standard discrete approach virtually for all ratings and are significantly less baised for mid ratings. The transition probability matrices also differ significantly in the overall mobility or dynamics that they imply. The discrete estimator provides matrices with a larger concentration of probability mass around the main diagonal. As the transition horizon increases, time homogeneities become apparent and the difference between the two hazard rate estimators is more marked. The heterogenous estimator emerges as more efficient and less biased for most ratings. The default probabilities from the homogeneous estimator are generally upward biased and larger overall migration mobility is implied. However, the discrepancies between homogeneous and heterogeneous estimators are less marked than those between discrete and continuous estimators. It turns out that the lower bound of 3bp recently established by the Basel Committee as the minimum transition probability for rare events is relatively conservative for the high credit-quality ratings. Another important implication, in the light of the New Basel Accord, is that the choice of either a discrete or a continuous framework for the estimation of sovereign default probabilities may result in substantially different levels of capital requirements.

There is evidence of non-Markov effects in the sovereign ratings which implies a specific type of heterogeneity in the rating process. Logit regression estimates suggest negative duration dependence for both downgrade and upgrade transition probabilities. Rating momentum effects are significant for downgrades (but not for upgrades) which is consistent with the rating agencies' practice of reducing a sovereign's credit-quality grade sequentially, notch by notch, rather than abruptly. These findings have important implications for risk management. For instance, in terms of pricing credit sensitive instruments it turns out that the rating momentum of a sovereign and its duration in the current rating may entail information about the future value of its debt obligations. In the case of multisovereign portfolios, upgrade and downgrade duration dependence for different assets is likely to cancel out on average so the effect might be less pronounced. This might not be the case for the momentum dependence because such effect is mostly present in downgrade migrations.

TABLE 5.1
One-year Homogeneous Rating Transition Probability Estimates

	Aaa	Aa	A	Baa	Ba	B	C	D
i) DM estimator: transition probabilities								
Aaa	0.94444444	0.05555556	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Aa	0.06395349	0.92441860	0.01162791	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
A	0.00000000	0.04123711	0.88659794	0.06185567	0.01030928	0.00000000	0.00000000	0.00000000
Baa	0.00000000	0.00000000	0.10000000	0.85454545	0.02727273	0.01818182	0.00000000	0.03333333
Ba	0.00000000	0.00000000	0.00000000	0.07575758	0.83333333	0.06818182	0.01515152	0.00757576
B	0.00000000	0.00000000	0.00000000	0.00000000	0.07142857	0.84523810	0.02380952	0.05952381
C	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.33333333	0.50000000	0.16666667
D	0	0	0	0	0	0	0	1
ii.1) HHR estimator: transition intensities								
Aaa	-0.06031064	0.06031064	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Aa	0.06345418	-0.07499131	0.01153712	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
A	0.00000000	0.03898427	-0.11695282	0.07796855	0.00000000	0.00000000	0.00000000	0.00000000
Baa	0.00000000	0.00000000	0.10085659	-0.15968960	0.05883301	0.00000000	0.00000000	0.00000000
Ba	0.00000000	0.00000000	0.00000000	0.08026789	-0.18972411	0.10215914	0.00000000	0.00729708
B	0.00000000	0.00000000	0.00000000	0.00000000	0.08042178	-0.19531004	0.06893296	0.04595530
C	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.30290456	-0.75726141	0.45435685
D	0	0	0	0	0	0	0	0
ii.2) HHR estimator: transition probabilities								
Aaa	0.94326537	0.05640615	0.00032019	0.00000816	0.00000012	0.00000000	0.00000000	**1.89×10^{-10}**
Aa	0.05934618	0.92974209	0.01050315	0.00040064	0.00000771	0.00000020	0.00000000	**0.00000002**
A	0.00113833	0.03549046	0.89328254	0.06804843	0.00196704	0.00006660	0.00000103	**0.00000556**
Baa	0.00003750	0.00175118	0.08802438	0.85780798	0.04957648	0.00252941	0.00005040	**0.00022258**
Ba	0.00000074	0.00004600	0.00347151	0.06763890	0.83258448	0.08550168	0.00244961	**0.00830707**
B	0.00000001	0.00000092	0.00009220	0.00270316	0.06680802	0.83799106	0.04368674	**0.04871789**
C	0.00000000	0.00000005	0.00000665	0.00025650	0.00935562	0.22501918	0.47564963	**0.28971236**
D	0	0	0	0	0	0	0	1

The bootstrap DGP parameters (transition probabilities in bold) are estimated from the observed 1981-2004 ratings. DM stands for discrete multinomial and HHR for homogenous hazard rate.

TABLE 5.2

Summary Statistics for Bootstrap 1-year Default Probability Estimates

	Panel A: DM estimator				
Rating	True PD	Mean($\widehat{PD}$)	StDev($\widehat{PD}$)	95% Conf. Int.	Bias
Aaa	1.9×10^{-10}	0	0	$[0,0]$	-1.89×10^{-10}
Aa	1.6×10^{-8}	0	0	$[0,0]$	-1.58×10^{-8}
A	5.6×10^{-6}	0	0	$[0,0]$	-5.56×10^{-6}
Baa	0.0002226	0.0002696	0.001502	$[0,0.00752]$	0.000047
Ba	0.008307	0.008905	0.008086	$[0,0.02801]$	0.000598
B	0.04872	0.04661	0.02596	$[0,0.1067]$	−0.00211
C	0.2897	0.2761	0.2591	$[0,1]$	−0.0136

	Panel B: HHR estimator				
Rating	True PD	Mean($\widehat{PD}$)	StDev($\widehat{PD}$)	95% Conf. Int.	Bias
Aaa	1.9×10^{-10}	2.3×10^{-10}	3.5×10^{-10}	$[0,1.2\times10^{-9}]$	3.5×10^{-11}
Aa	1.6×10^{-8}	1.8×10^{-8}	2.7×10^{-8}	$[0,9\times10^{-8}]$	2.6×10^{-9}
A	5.6×10^{-6}	6.6×10^{-6}	6.8×10^{-6}	$[1.8\times10^{-7},2.4\times10^{-5}]$	9.9×10^{-7}
Baa	0.0002226	0.0002701	0.0002625	$[1.1\times10^{-5},0.0009718]$	0.0000475
Ba	0.008307	0.01003	0.007823	$[0.0007534,0.02963]$	0.00173
B	0.04872	0.05861	0.04568	$[0.01333,0.1204]$	0.00989
C	0.2897	0.3220	0.1882	$[0,0.7488]$	0.0323

See footnote to Table 5.1. True PD are the 1-year default probability parameters in the bootstrap DGP for the ratings. The bias is calculated as $Mean(\widehat{PD}) - True(PD)$.

TABLE 5.3

Bootstrap Tests for Differences between Discrete and Continuous Estimator

Panel A: 1-year default probability differential				
Rating	Mean$\Delta(\widehat{PD})$	StDev$\Delta(\widehat{PD})$	95% Conf. Int.	Null $\Delta(PD) = 0$
Aaa	2.3×10^{-10}	3.5×10^{-10}	$[0, 1.2 \times 10^{-9}]$	Not reject
Aa	1.8×10^{-8}	2.7×10^{-8}	$[0, 9 \times 10^{-8}]$	Not reject
A	6.6×10^{-6}	6.8×10^{-6}	$[1.8 \times 10^{-7}, 2.4 \times 10^{-5}]$	Reject
Baa	4.8×10^{-7}	0.001486	[−0.006857, 0.000946]	Not reject
Ba	0.001129	0.004858	[−0.01078, 0.01024]	Not reject
B	0.01200	0.03948	[−0.01468, 0.04364]	Not reject
C	0.04557	0.2171	[−0.3796, 0.5269]	Not reject

Panel B: Matrix mobility differential

$\hat{m}$	$\Delta\hat{m}$	Mean($\Delta\hat{m}$)	StDev($\Delta\hat{m}$)	95% Conf. Int.	$\Delta m = 0$
0.1921(HHR) 0.1954(DM)	−0.003257	−0.09424	0.04160	[−0.1801, −0.03485]	Reject

See note in Table 5.1. The default probability differential, $\Delta(PD)$, and the mobility differential, $\Delta\hat{m}$, are computed as the HHR minus the DM estimator.

TABLE 5.4

Longer Horizon Homogeneous Transition Probability Matrices

Panel A: 2-year time-homogeneous transition matrices

	Aaa	Aa	A	Baa	Ba	B	C	D
i) DM estimator: transition probabilities								
Aaa	0.89393939	0.10606061	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Aa	0.13253012	0.84337349	0.02409639	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
A	0.00000000	0.09302326	0.83720930	0.04651163	0.02325581	0.00000000	0.00000000	0.00000000
Baa	0.00000000	0.00000000	0.19230769	0.73076923	0.03846154	0.01923077	0.00000000	0.01923077
Ba	0.00000000	0.00000000	0.01587302	0.12698413	0.73015873	0.11111111	0.01587302	0.00000000
B	0.00000000	0.00000000	0.00000000	0.02564103	0.12820513	0.71794872	0.02564103	0.10256410
C	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	1	0.00000000	0.00000000
D	0	0	0	0	0	0	0	1
ii) HHR estimator: transition probabilities								
Aaa	0.89309742	0.10566051	0.00118122	0.00005910	0.00000169	0.00000006	0.000000002	0.00000001
Aa	0.11116780	0.86814133	0.01920177	0.00143194	0.00005410	0.00000274	0.00000005	0.00000025
A	0.00419941	0.06488333	0.80432402	0.11930575	0.00677385	0.00045614	0.00001340	0.00004410
Baa	0.00027174	0.00625894	0.15432853	0.74518679	0.08414610	0.00858763	0.00029323	0.00092705
Ba	0.00001060	0.00032267	0.01195477	0.11480340	0.70235638	0.14498126	0.00706058	0.01851035
B	0.00000032	0.00001280	0.00062900	0.00913344	0.11247287	0.72636960	0.05735495	0.09402703
C	0.00000004	0.00000125	0.00008910	0.00162146	0.02943731	0.34594667	0.24106727	0.38183694
D	0	0	0	0	0	0	0	1

Panel B: 3-year time-homogeneous transition matrices

	Aaa	Aa	A	Baa	Ba	B	C	D
i) DM estimator: transition probabilities								
Aaa	0.825	0.175	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Aa	0.2	0.76363636	0.03636364	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
A	0.00000000	0.15384615	0.73076923	0.07692308	0.03846154	0.00000000	0.00000000	0.00000000
Baa	0.00000000	0.00000000	0.23333333	0.63333333	0.06666667	0.033333	0.00000000	0.03333333
Ba	0.00000000	0.00000000	0.05	0.15000000	0.6	0.175	0.025	0.00000000
B	0.00000000	0.00000000	0.00000000	0.05263158	0.10526316	0.78947368	0.05263158	0.00000000
C	0.00000000	0.00000000	0.00000000	0.00000000	1	0.00000000	0.00000000	0.00000000
D	0	0	0	0	0	0	0	1
ii) HHR estimator: transition probabilities								
Aaa	0.84869989	0.14865507	0.00245627	0.00018061	0.00000769	0.0000004	0.00000002	0.00000004
Aa	0.15640335	0.81410236	0.02643214	0.00288804	0.00016023	0.00001240	0.00000020	0.00000130
A	0.00873241	0.08931496	0.72970394	0.15754915	0.01317806	0.00131647	0.00006220	0.00014276
Baa	0.00083059	0.01262352	0.20379847	0.65546943	0.10786059	0.01656732	0.00065266	0.00219742
Ba	0.00004820	0.00095553	0.02325718	0.14715790	0.60046186	0.18628034	0.01267560	0.02916336
B	0.00000196	0.00005770	0.00180806	0.01754222	0.14351949	0.64937460	0.05295081	0.13474512
C	0.00000049	0.00000641	0.00040229	0.00419626	0.05450719	0.39994490	0.18600557	0.35493687
D	0	0	0	0	0	0	0	1

(.cont)

Panel C: Matrix mobility metric for varying horizon			
Transition horizon (years)	$\hat{m}$(HHR)	$\hat{m}$(DM)	$\Delta\hat{m}$
1	0.1921	0.1954	-0.0033
2	0.3155	0.3916	-0.0761
3	0.3802	0.4661	-0.0859
4	0.4007	0.5197	-0.1190

DM stands for discrete multinomial and HHR for homogenous hazard rate.
The mobility differential, $\Delta\hat{m}$, is computed as the HHR minus the DM estimator.

TABLE 5.5

Longer Horizon Heterogeneous Transition Probability Matrices

Panel A: NHHR transition probabilities								
	Aaa	Aa	A	Baa	Ba	B	C	D
Two-year horizon								
Aaa	0.92016251	0.07983749	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Aa	0.17143835	0.80310470	0.02545695	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
A	0.00336537	0.06106579	0.78600696	0.13813270	0.01075966	0.00066952	0.00000000	0.00000000
Baa	0.00000000	0.00187119	0.15640571	0.71752941	0.11918620	0.00488992	0.00005879	0.00005879
Ba	0.00000000	0.00006791	0.01368107	0.10337922	0.72815887	0.13087669	0.00475398	0.01908226
B	0.00000000	0.00000000	0.00015967	0.00525031	0.11132463	0.77381518	0.02885281	0.08059740
C	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.16111111	0.67063492	0.16825397
D	0	0	0	0	0	0	0	1
Three-year horizon								
Aaa	0.89704219	0.10295781	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
Aa	0.24791551	0.71512515	0.03695934	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
A	0.00477332	0.08545516	0.75038362	0.14407716	0.01308519	0.00162502	0.00002053	0.00058000
Baa	0.00000000	0.00271339	0.22693730	0.58257029	0.17611914	0.00949798	0.00016421	0.00199768
Ba	0.00000000	0.00028635	0.02743720	0.12277370	0.62134554	0.19443597	0.00740273	0.02631851
B	0.00000000	0.00005206	0.00412252	0.01114124	0.13311615	0.70687773	0.03902732	0.10566297
C	0.00000000	0.00000000	0.00017013	0.00221167	0.04689865	0.19957862	0.50550336	0.24563758
D	0	0	0	0	0	0	0	0

Panel B: Matrix mobility

Transition horizon (years)	$\hat{m}$(HHR)	$\hat{m}$(NHHR)	$\Delta\hat{m}$
1	0.1921	0.1303	0.0618
2	0.3155	0.2399	0.0755
3	0.3802	0.3146	0.0786
4	0.4007	0.4210	-0.0204

HHR stands for homogenous hazard rate (see probability estimates in Table 5.4) and NHHR stands for non-homogeneous hazard rate estimator. The mobility differential, $\Delta\hat{m}$, is computed as the HHR minus the NHHR estimator.

TABLE 5.6

Summary Statistics for Bootstrap 2-year Default Probability Estimates

Panel A: HHR estimator

Rating	True PD	Mean($\widehat{PD}$)	StDev($\widehat{PD}$)	95% Conf. Int.	Bias
Aaa	0	7.7×10^{-9}	1.1×10^{-8}	$[0, 4 \times 10^{-8}]$	7.7×10^{-9}
Aa	0	3.3×10^{-7}	4.4×10^{-7}	$[0, 1.6 \times 10^{-6}]$	3.3×10^{-7}
A	4.8×10^{-6}	5.6×10^{-5}	5.3×10^{-5}	$[3.1 \times 10^{-6}, 0.0001931]$	5.1×10^{-5}
Baa	1.2×10^{-4}	0.0011894	0.0009514	[0.0001018, 0.003780]	0.001069
Ba	0.02519	0.0235	0.01371	[0.003997, 0.05473]	−0.00169
B	0.09319	0.1221	0.04872	[0.04018, 0.2360]	0.02895
C	0.2232	0.4684	0.2123	[0, 0.8396]	0.2452

Panel B: NHHR estimator

Rating	True PD	Mean($\widehat{PD}$)	StDev($\widehat{PD}$)	95% Conf. Int.	Bias
Aaa	0	0	0	[0, 0]	0
Aa	0	0	0	[0, 0]	0
A	4.8×10^{-6}	3.4×10^{-6}	1.5×10^{-5}	$[0, 4.7 \times 10^{-5}]$	-1.3×10^{-6}
Baa	1.2×10^{-4}	0.0001053	0.0002146	[0, 0.0007048]	−0.00001486
Ba	0.02519	0.02406	0.01417	[0.002985, 0.05641]	−0.001134
B	0.09319	0.08875	0.03324	[0.02757, 0.1603]	−0.0044
C	0.2232	0.18249	0.07407	[0.04106, 0.3522]	−0.0407

The bootstrap DGP parameters (true transition probabilities) are estimated from the observed 1981-2004 ratings. The bias is calculated as Mean($\widehat{PD}$)−True(PD). HHR stands for homogeneous hazard rate and NHHR for non-homogeneous hazard rate estimator.

TABLE 5.7

Bootstrap Tests for Differences between Homogeneous and Heterogeneous Estimator

	Panel A: 2-year default probability differential			
Rating	Mean$\Delta(\widehat{PD})$	StDev$\Delta(\widehat{PD})$	95% Conf. Int.	Null $\Delta(PD)=0$
Aaa	7.7×10^{-9}	1.1×10^{-8}	$[0, 4\times10^{-8}]$	Not reject
Aa	3.3×10^{-7}	4.4×10^{-7}	$[0, 1.6\times10^{-6}]$	Not reject
A	5.2×10^{-5}	5.5×10^{-5}	$[-1.1\times10^{-5}, 0.0001924]$	Not reject
Baa	0.001084	0.0009477	$[-3.8\times10^{-5}, 0.003562]$	Not reject
Ba	-0.0005521	0.00898	$[-0.01936, 0.01734]$	Not reject
B	0.03339	0.03171	$[-0.03262, 0.10174]$	Not reject
C	0.2859	0.1963	$[-0.1717, 0.6609]$	Not reject

	Matrix mobility metric differentials				
$\hat{m}$	$\Delta\hat{m}$	Mean$(\Delta\hat{m})$	StDev$(\Delta\hat{m})$	95% Conf. Int.	$\Delta m=0$
0.3155(HHR)	0.07555	0.08597	0.0364	$[0.0134, 0.1540]$	Reject
0.2399(NHHR)					

See note in Table 5.6. The differential statistics, $\Delta(\widehat{PD})$ and $\Delta\hat{m}$, are computed as the HHR minus the NHHR value.

TABLE 5.8
Logit Test Results for Non-Markov Effects

Panel A: Duration effect

	Upgrade		Downgrade	
Coefficient	estimate	t ratio	estimate	t ratio
β	-1.88	-5.33	-1.65	-5.74
γ_{1998}	1.99	2.71	-0.31	-0.37
γ_{1999}	0.48	0.39	1.68	3.12
γ_{2000}	1.36	1.69	1.49	2.62
γ_{2001}	3.81	4.88	0.37	0.43
γ_{2002}	2.39	2.89	1.94	2.70
γ_{2003}	3.78	5.67	-0.08	-0.12
γ_{2004}	-0.44	-0.67	-0.90	-1.66
α	-1.57	-1.50	-0.41	-1.15
ξ	0.27	1.04	0.11	0.44

Panel B: Momentum effect

	Upgrade		Downgrade	
Coefficient	estimate	t ratio	estimate	t ratio
β	-1.24	-1.02	1.20	2.94
γ_{1998}	1.65	2.75	-0.37	-0.46
γ_{1999}	0.60	0.71	1.21	2.67
γ_{2000}	0.84	1.19	0.71	1.47
γ_{2001}	1.83	3.23	-0.79	-0.99
γ_{2002}	0.84	1.16	0.35	0.63
γ_{2003}	2.61	5.47	-0.28	-0.45
γ_{2004}	0.90	1.08	-0.08	-0.12
α	-6.37	-14.46	-5.18	-14.06
ξ	-0.027	-0.35	-0.17	-2.10

The logit estimates are based on monthly ratings March, 1981-March 2004. The logits in Panel A and B model the probability of upgrade/downgrade as a function of duration and rating drift (β), respectively. γ_j are dummies, α is an intercept and ξ is a parameter in the disturbance variance equation to capture groupwise heteroskedasticity (industrial/non-industrial sovereigns).

Figure 5.1: Aggregate Ratings Distribution, 1981-2003 (Moody's)

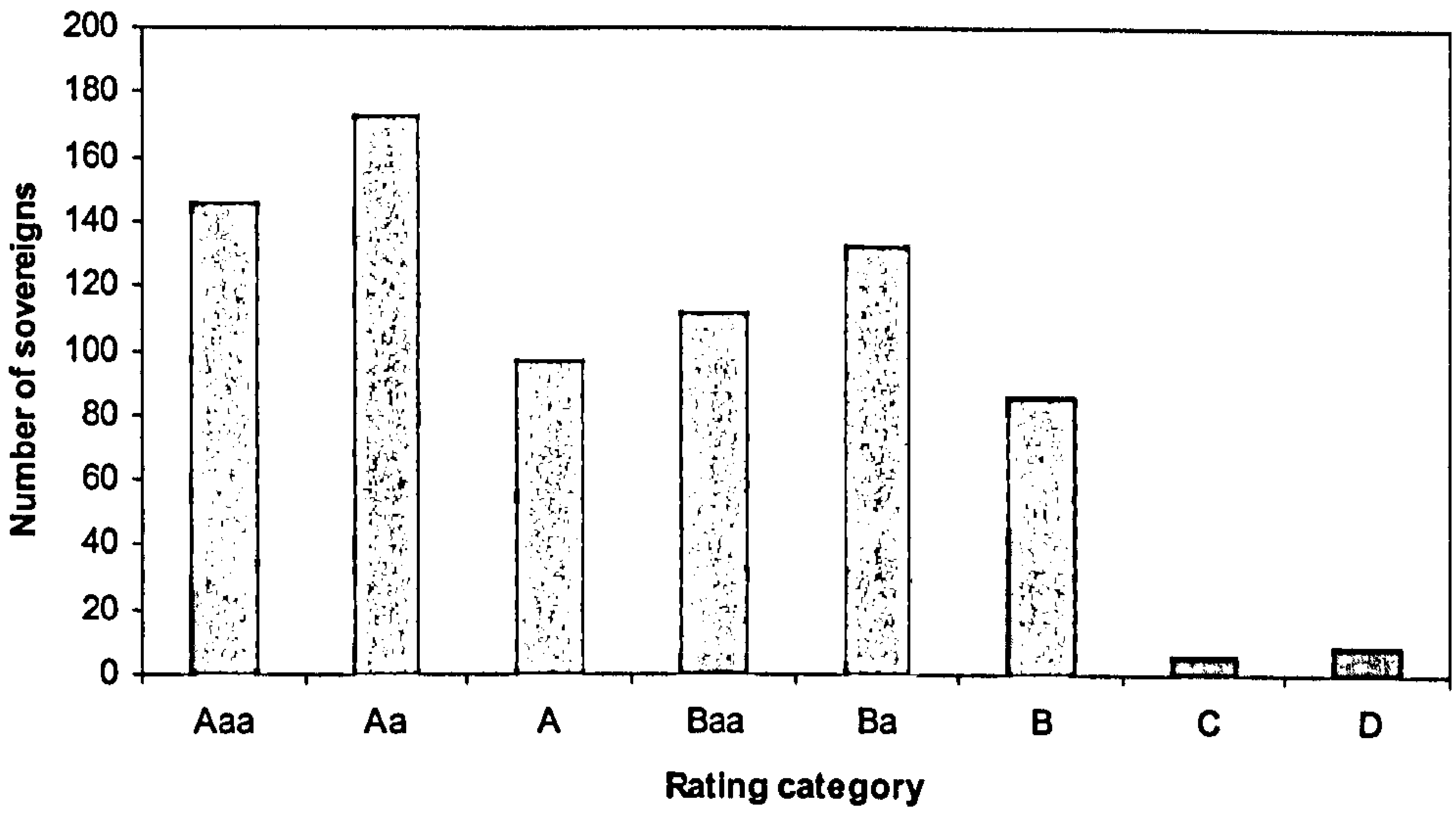
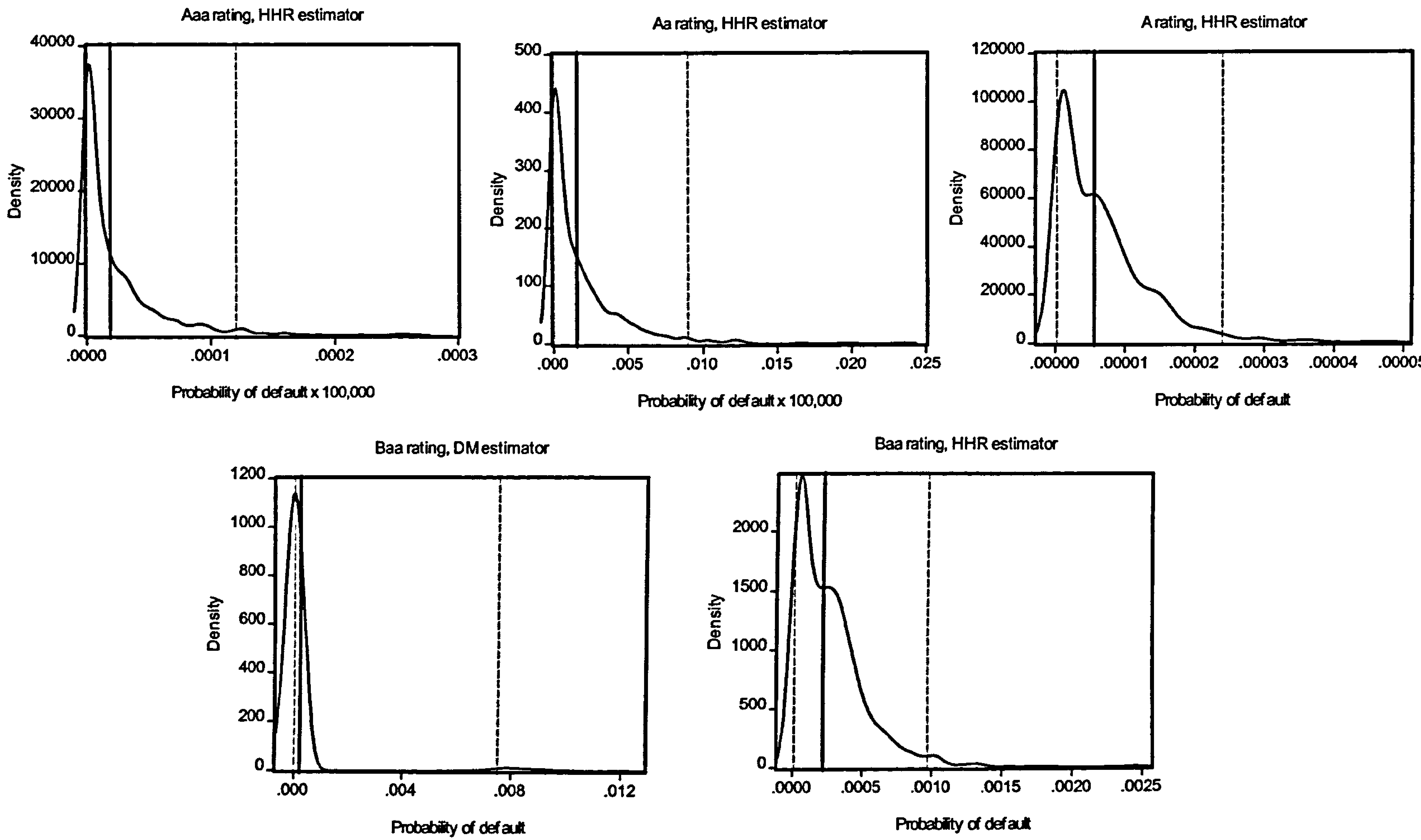
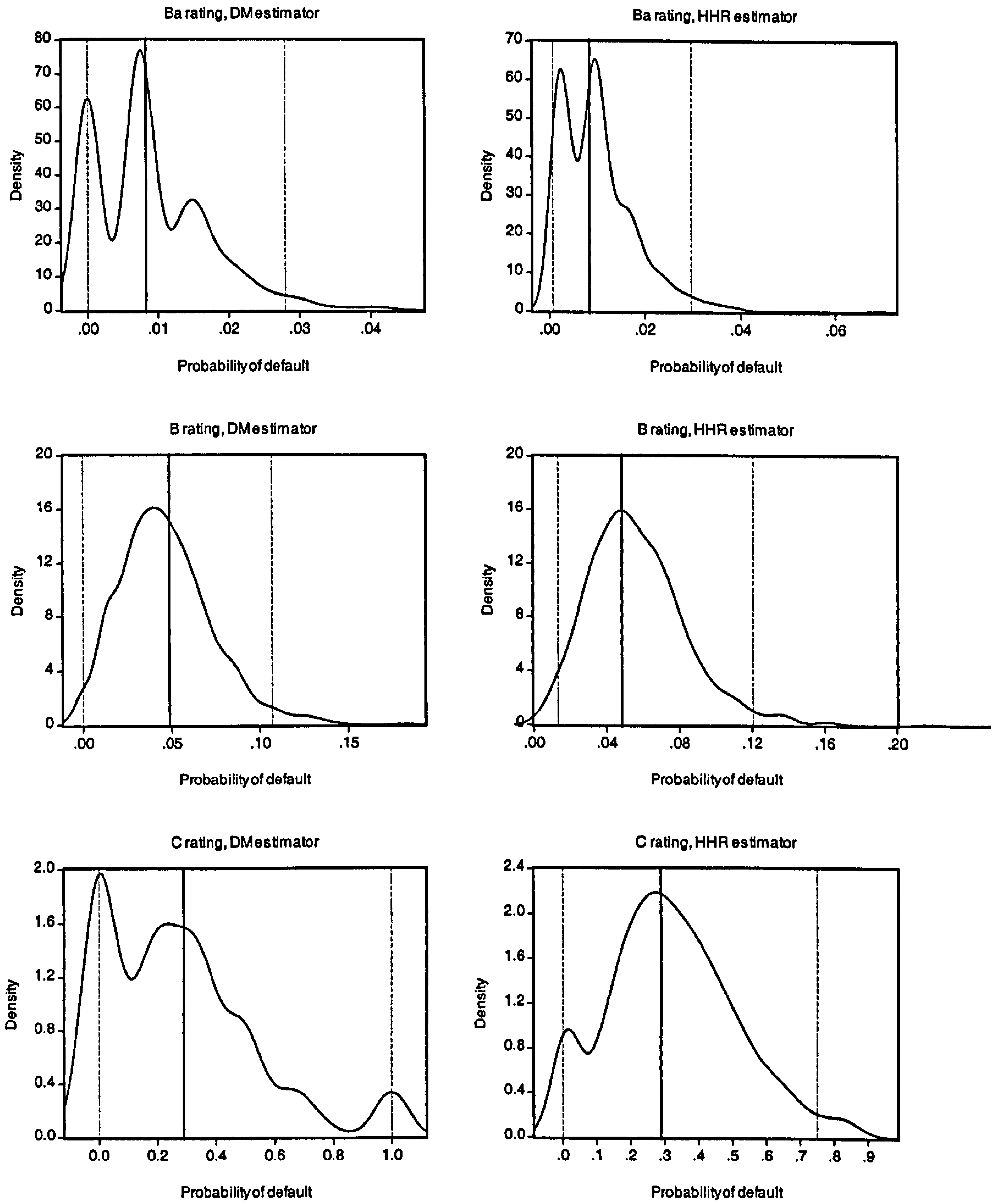


Figure 5.2: Discrete versus Continuous Bootstrap Default Probability Estimates
(high credit-quality ratings)



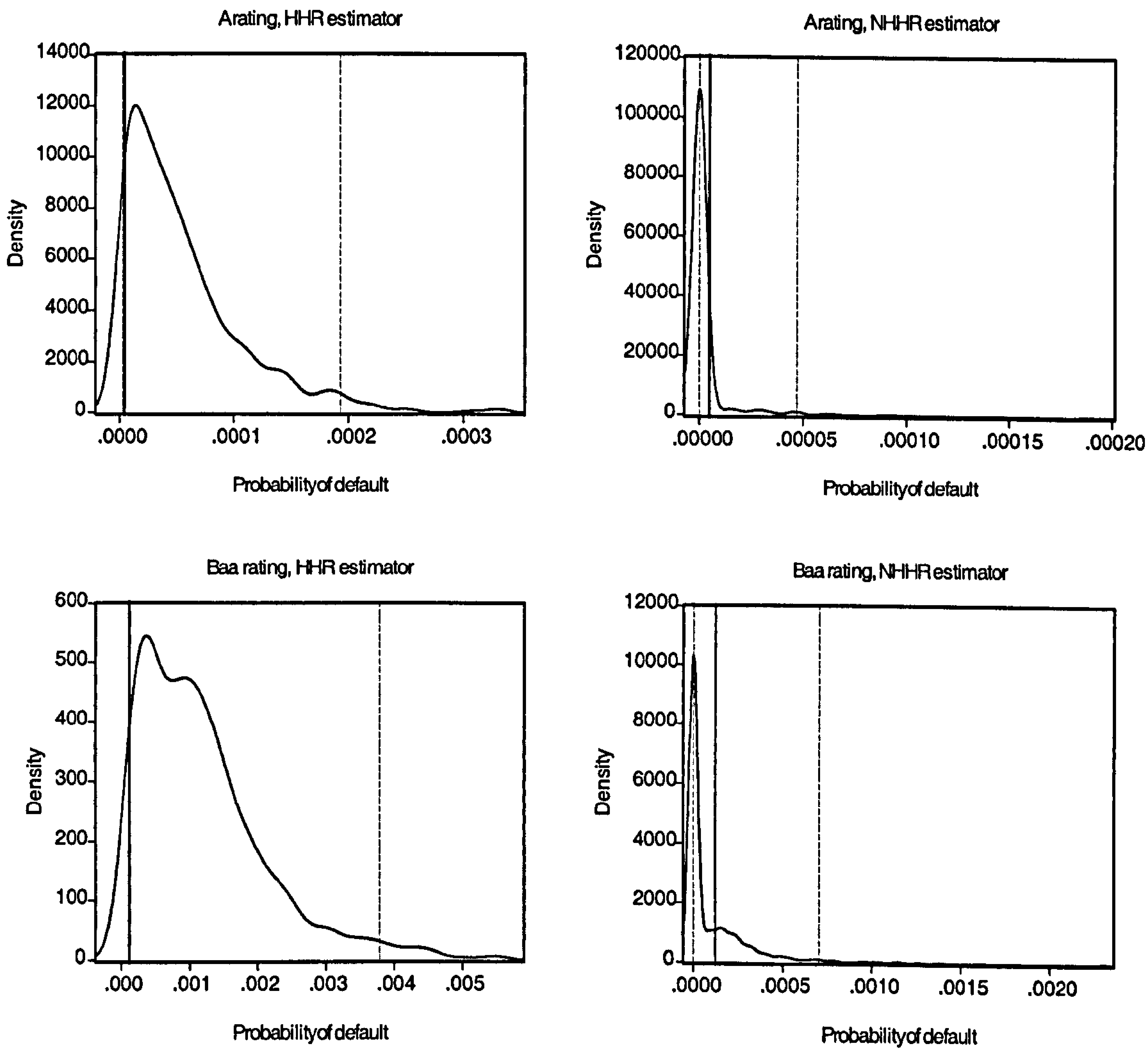
Note: The plotted densities are based on the Gaussian kernel function. The bold line represents the "true" one-year default probability. The dotted lines correspond to the 95% confidence intervals. DM is the discrete multinomial estimator and HHR is the continuous homogeneous hazard rate estimator.

Figure 5.3: Discrete versus Continuous Bootstrap Default Probability Estimates
(low credit-quality ratings)



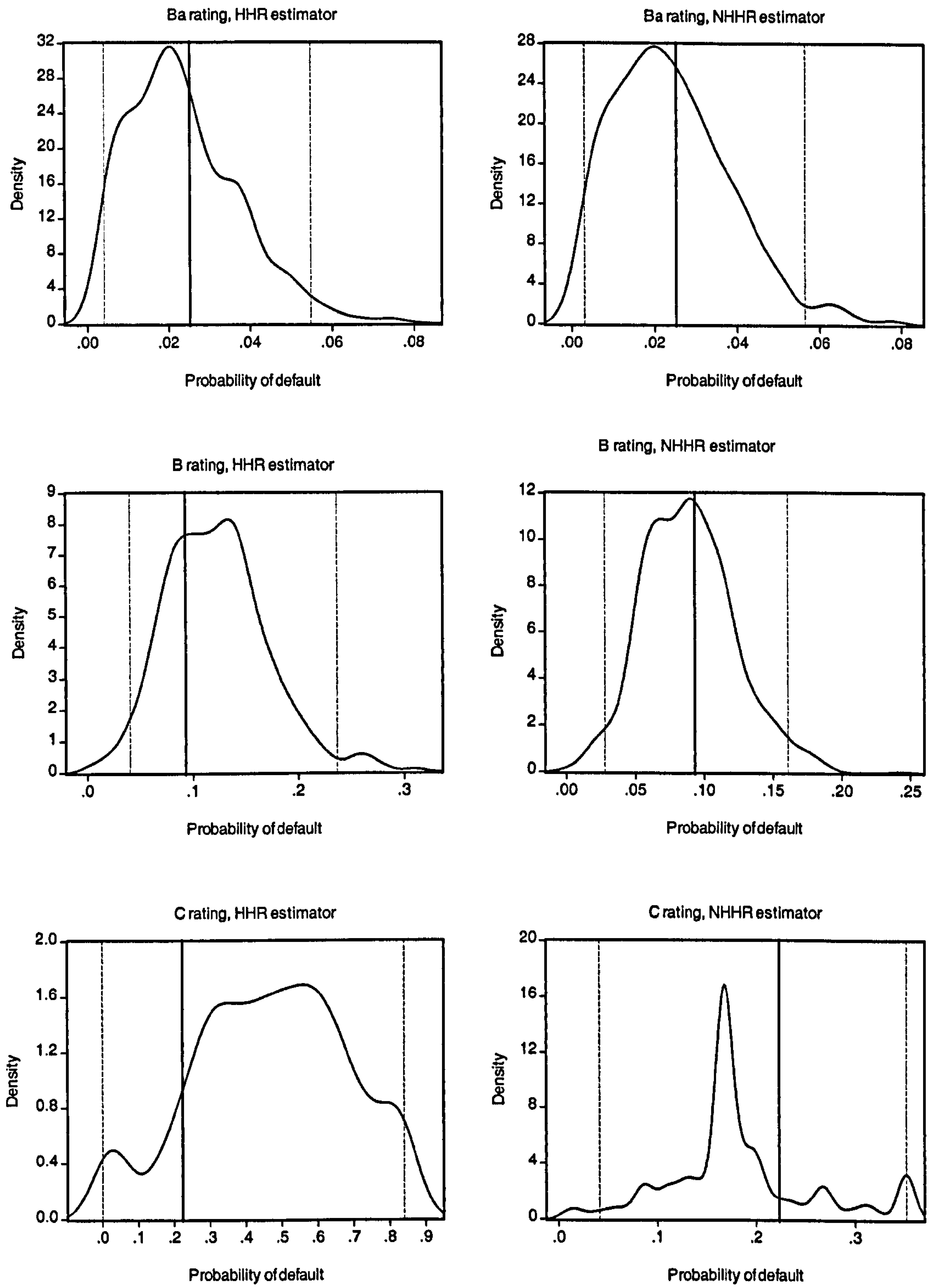
Note: The plotted densities are based on the Gaussian kernel function. The bold line represents the "true" one-year default probability. The dotted lines correspond to the 95% confidence intervals. DM is the discrete multinomial estimator and HHR is the continuous homogeneous hazard rate estimator.

Figure 5.4: Homogeneous versus Heterogeneous Bootstrap Default Probability Estimates
(high credit-quality ratings)



Note: The plotted densities are based on the Gaussian kernel function. The bold line represents the "true" two-year default probability. The dotted lines correspond to the 95% confidence intervals. HHR is the continuous homogeneous hazard rate estimator and NHHR is the continuous heterogeneous hazard rate estimator.

Figure 5.5: Homogeneous versus Heterogeneous Bootstrap Default Probability Estimates
(low credit-quality ratings)



Note: The plotted densities are based on the Gaussian kernel function. The bold line represents the "true" two-year default probability. The dotted lines correspond to the 95% confidence intervals. HHR is the continuous homogeneous hazard rate estimator and NHHR is the continuous heterogeneous hazard rate estimator.

Figure 5.6: Eigenvalues and Transition Horizon

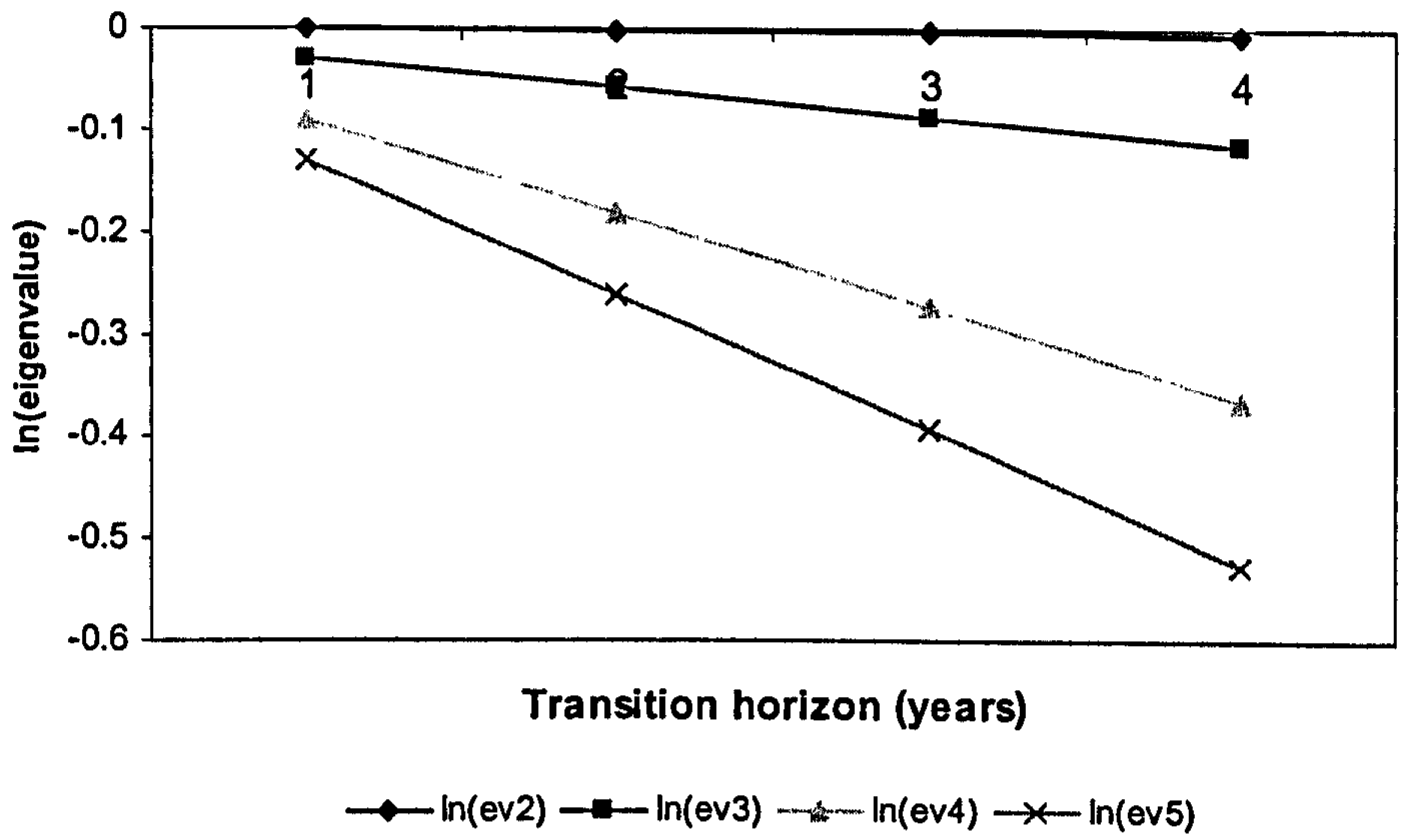
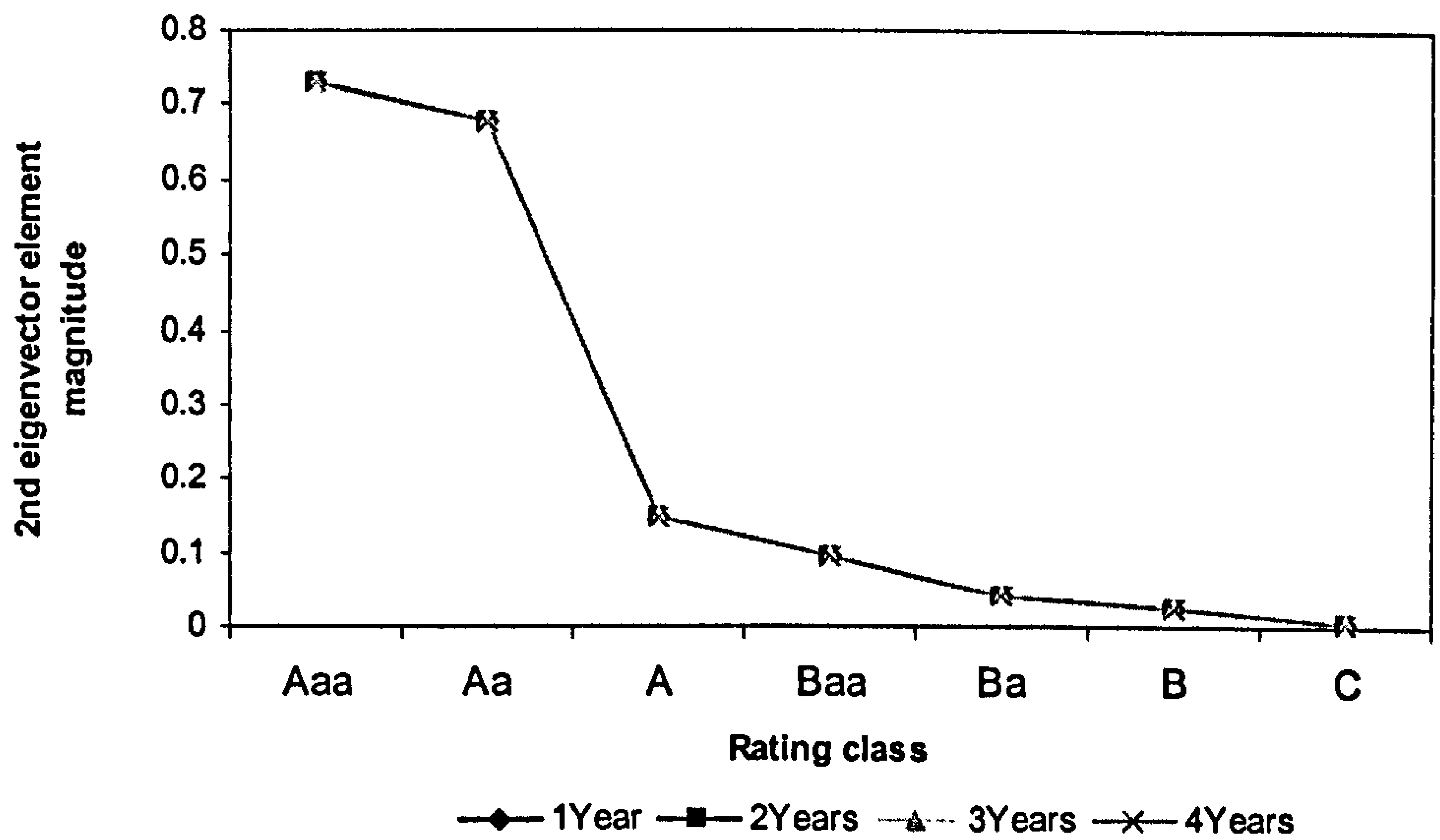


Figure 5.7: Second Eigenvector for Different Transition Horizons



CHAPTER 6: CONCLUSION

6.1 Concluding Remarks

It is important to recognize the complexity of international debt crises in the sense that they are determined by the interaction of economic fundamentals and a range of unobserved factors — political, social, cultural and legal attributes could equally determine the timely fulfillment of debt obligations. The development of better methods to anticipate sovereign default and, more broadly, to estimate rating migration probabilities falls under the umbrella of sovereign risk analysis. The main stakeholders are the policymakers of emerging and developing markets, the financial institutions and the bank regulators.

We start by investigating the importance of accounting for heterogeneity in panel sovereign default models. The literature has developed a critical attitude towards the restrictive country-homogeneity assumption in panel sovereign default models. However, direct evidence on the importance of relaxing the latter, particularly for forecasting purposes, has not yet been provided. Chapter 2 seeks to complement the literature in this respect. Several panel logit specifications are considered ranging from a simple pooled model to a time-varying random coefficients model.

Statistical hypotheses tests and model selection criteria strongly suggest that heterogeneous models better describe the data for our large sample of emerging markets and LDCs. The different models are, however, similar regarding the plausibility of the coefficient signs. In order to develop an Early Warning System (EWS) for sovereign default it seems natural to compare these models on the basis of their forecast performance. To this end, we conduct a comprehensive forecast evaluation exercise which fills a gap in the panel sovereign default literature. Our out-of-sample, recursive forecast framework comprises statistical and economic loss functions as well as formal tests of predictive performance both across models and relative to naive forecasts. As expected, model ranking depends on the decision-maker's loss function. The analysis reveals that logits that exclusively control for either time or regional heterogeneity in a simple manner forecast relatively well. Fully heterogeneous models produce poorer forecasts. Our results in Chapter 2 provide further evidence, which is new in the context of sovereign default, that the model that best describes the data is not necessarily good in terms of forecasting.

Little attention has been paid to residual autocorrelation in panel logit (or probit) models of sovereign default. This stems both from the overlapping problem inherent in EWSs and from the serial dependence in the macroeconomic and financial ratios typically used as regressors. The lack of attention can perhaps be attributed to the fact that the usual econometric packages do not produce an autocorrelation-robust estimator for panel logit and probit modeling. Neglecting residual autocorrelation will invalidate inferences on the determinants of default because the usual standard errors are biased.

In Chapter 3 we demonstrate the magnitude of the serial correlation problem in the context of several panel logit models of sovereign default. Building on the correction proposed by Estrella and Rodriguez (1998) for panel probits, we implement an autocorrelation-robust estimator for panel logits along the lines of the Newey-West approach. The results suggest that the usual ML standard errors are substantially underestimated. Inferences based on them are overturned, for several regressors, when the robust standard errors are used. Our findings in this regard have broad implications since the same criticism applies to much of the recent literature on EWSs for currency/banking crises which has largely utilized panel binary-choice models.

Although most of the extant EWSs are based on panel binary-choice models for macroeconomic and financial indicators, credit ratings have also been considered. To the extent that they capture unobservable sociopolitical factors as well as market expectations, there may be gains from combining them with forecasts from macro/financial logits. Nonparametric classification techniques allow researchers to avoid the problem of having to specify a functional form and thus, may better capture complex relationships. Chapter 4 investigates several issues in the development of an optimal EWS for sovereign default using three classifiers — a logit regression and K-means clustering, both based on macrovariables, and a logit regression based on bankers' credit ratings.

We demonstrate the potential importance of optimally calibrating these classifiers according to the decision-maker's preferences. In particular, it is shown that the optimal warning horizon, cut-off probability, assignment rule and number of clusters vary with the loss function. These parameters have mostly been chosen in an ad hoc manner in the literature. We corroborate the different strengths of the classifiers in terms of Type I and Type II errors and further demonstrate that there are out-of-sample forecast gains from combining them. Finally, evidence is provided on the importance of accounting for the decision-maker's loss function in choosing the forecast combining scheme also. We believe that important practical recommendations emerge from Chapter 4 which should also be relevant for the development of EWSs for currency and banking crises.

Having analysed the above issues regarding the prediction of sovereign default events, we next turn to the estimation of rating transition probability matrices. The small histories of sovereign ratings (and transitions) available pose problems for the reliability of sovereign transition matrix estimates. In this context, it is important to investigate the small-sample properties of rival estimators. Moreover, most of the estimators available build on the notion that the current transition probability does not depend on the ratings history (Markov property). Thus, it is relevant to test for non-Markov effects such as momentum and duration dependence. There is no evidence on these issues in the context of sovereign rating migrations and this thesis attempts to fill this gap. For this purpose, in Chapter 5 we compare the bias, efficiency and implied overall mobility of continuous versus discrete estimators, on the one hand, and of time-homogeneous versus time-heterogeneous estimators on the other. The comparison is based on a bootstrap method that facilitates their empirical distribution.

Two interesting findings emerge. First, the accuracy of transition probability estimators is enhanced when we move from a discrete to a continuous-time framework that utilizes rating duration information. This result is important given that the discrete (multinomial) approach is the standard method used by leading rating agencies. The average migration dynamics implied by discrete and continuous estimators is also significantly different as measured by a matrix mobility metric based on the spectral decomposition. Second, continuous estimators that account for time-heterogeneity are generally more efficient and less biased. Finally, we document duration dependence for both sovereign downgrades and upgrades which implies that transition intensity significantly increases with duration. Rating momentum effects are prominent only for downgrades. This is consistent with the common practice by agencies of reducing a sovereign's credit-quality grade gradually. Overall our results show that several stylised facts of corporate credit ratings also apply to sovereigns.

6.2 Further Research

Political risk indexes have been recently constructed (PRS-Political Risk Services) that incorporate the degree of corruption, ethnic tensions, democracy, military involvement in politics as well as the socioeconomic and investment profile of countries. Lack of publicly available data has precluded the use of such factors in our analysis. Examination of the latter as a complementary tool for sovereign risk assessment might be of interest.

Our results have suggested that controlling for regional heterogeneity is important in the context of sovereign default prediction. A natural follow-up would be to investigate the existence of unsystematic region-specific risk factors over and above the underlying common factors driving world debt repayment performance. For example, given that Latin America has experience a larger degree of liberalization compared to Asia factors reflecting global trade links and foreign direct investment may play a more important role in future debt repayment. The regional differences may thus materialize in terms of picking up different default determinants and would likely increase the predictive performance of sovereign default events by region. Moreover, given the pervasive evidence regarding the magnitude of the serial correlation problem inherent in financial crisis EWSs, we could further support the proposed correction by means of Monte Carlo simulations. By simulating the macro variables as autoregressive processes and the debt crisis variable using logistic error terms, we can show whether the correction provides accurate standard errors. The latter can be facilitated by comparing the variance of the parameter estimates across simulations with the usual and autocorrelation corrected standard errors.

In addition, the period between 1994 to date deserves particular attention given the general turbulence in the sovereign bond market. The latter has been characterized by large defaults on foreign currency bonds. By contrast, most of the default episodes in the pre-1994 era are attributed to defaults on bank syndicated loans and trade credit obligations. Focusing the analysis exclusively on emerging markets with access to the sovereign bond market could provide useful insights. However, given the scarcity of bond default events in this context, a broader definition of 'sovereign distress' is needed that reflects increased credit risk but not necessarily default. In parallel with the distressed debt literature in corporate finance one can define the occurrence of a 'sovereign distress' event when the spreads of the most liquid bonds exceed a threshold value of, say 1000 basis points. The availability of actively traded bonds for these sovereigns enables the comparison of different classification techniques on the basis of economic significance, that is, one could simulate simple trading strategies and return distributions of emerging market sovereign bond portfolios. For example, recursively estimating binary or multinomial logit models using rolling windows of data, out-of-sample forecasts for the dynamics of the risk-neutral implied default probabilities can be generated. The ability of trading strategies based on out-of-sample default probability forecasts to beat the 'buy & hold' strategy in terms of emerging market bond portfolio returns could be assessed. A naive trading strategy could be constructed as follows: increase the position upon a clear upgrading signal, close the position upon a clear downgrading signal and keep the position unchanged upon no clear signal, given a minimum investment required to re-open a position and borrowing and lending at the risk-free rate.

Finally, the current comparative analysis of estimators for sovereign rating migration matrices could be extended by addressing whether the differences uncovered translate into economic significance. For this purpose, once could test whether the estimators lead to different VaR capital requirements for sovereign credit portfolios. More precisely, a sovereign bond portfolio that mimics the distribution of the Moody's sovereign rating universe on a specific date can be constructed. Industrial credit portfolio models can be used to generate the value distribution of the portfolio of bonds, where besides recovery rates and credit spreads, the crucial input parameter will be the credit rating migration matrix from the relevant estimator. The question would be: what is the *ceteris-paribus* portfolio value distribution one year ahead using different transition matrices? The simulated portfolio value distribution will provide the risk levels and the 99% standard VaR economic risk capital requirements. The discrepancy surrounding the different methods to calibrate the matrices may thus be assessed in term of economic relevance. If the differences between estimators are economically significant, their implied VaR capital levels will exhibit sizeable differences.

References

[1] Aalen, O.O. and S. Johansen (1978) An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5, 141-150.

[2] Abiad, A. (2003) Early warning systems: A survey and a regime-switching approach. *IMF Working Paper* 03/32.

[3] Aiolfi, M., and A. Timmermann (2003). Persistence in forecasting performance. *Mimeo,* University of California at San Diego.

[4] Albanis, G. and R.A. Batchelor (1999) Combining heterogeneous classifiers for stock selection, *Mimeo*, Cass Business School, London.

[5] Ali, K.M., and M.J. Pazzani (1995) Error reduction through learning multiple descriptions, *Machine Learning,* 24, 173-2002.

[6] Alpayadin, E. (1998). Techniques for combining multiple learners, *Proceedings of Engineering of Intelligent Systems Conference,* 2, 6-12.

[7] Altman, E.I., H. Frydman and D.L. Kao (1985) Introducing recursive partitioning for financial classification: the case of financial distress. *Journal of Finance*, XL, 269-291.

[8] Altman, E.I. (1998) The importance and subtlety of credit rating migration. *Journal of Banking and Finance*, 22, 1231-1247.

[9] Altman, E.I. and A. Saunders (1997) Credit risk measurement: developments over the last twenty years. *Journal of Banking and Finance*, 21, 1721-1742.

[10] Altman, E.I. and D.L. Kao (1992) The implications of corporate bond ratimg drift. *Financial Analysts Journal*, May/June, 64-75.

[11] Arellano, M. and B. Honore (2001) Panel data models: Some recent developments. In *Handbook of Econometrics*, 5, edited by J.J.Heckman and E.Leamer, Amsterdam, North Holland.

[12] Arora, V. and M. Cerisola (2001) How does US monetary policy influence sovereign spreads in emerging markets? IMF Staff Papers 48, 474-498.

[13] Arvanitis, A., J. Gregory and J-P. Laurent (1999) Building models for credit spreads. *Journal of Derivatives*, Spring, 27-43.

[14] Avramovic, D. (1958) *Debt servicing capacity and post-war growth in international indebtness*, John Hopkins Press, Baltimore, Maryland.

[15] Aylward, L. and R. Thorne (1998) An econometric analysis of coutries' repayment performance to the IMF. *IMF Staff Paper 45.*.

[16] Babbel, D.F. (1996) Insuring sovereign debt against default, World Bank Discussion Paper, 328, Washington: The World Bank.

[17] Bahar, R. and K. Nagpal (2000) Rating transitions: modelling the dynamics. *Credit*, March, 57-63.

[18] Balkan, E.M. (1992) Political instability, country risk and probability of default. *Applied Economics*, 24, 99-1008.

[19] Baltagi, B. H. (2002) *Econometric Analysis of Panel Data*, 2nd Edition. New York: John Wiley.

[20] Baltagi, B. H. and J. M. Griffin (1997) Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, 77, 303-327.

[21] Baltagi, B. H., J. M. Griffin and W. Xiong (2000) To Pool or not to Pool: Homogeneous versus Heterogeneous Estimators Applied to Cigarette Demand. *The Review of Economics and Statistics*, 82(1), 117-126.

[22] Baltagi, B.H. and Q. Li (1995) Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics* 68, 133-151.

[23] Bangia, A., F.X. Diebold, A. Kronimus, C. Schagen and T. Schuermann (2002) Ratings migration and the business cycle, with applications to portfolio stress testing. *Journal of Banking and Finance*, 26, 445-474.

[24] Basel Committee on Banking Supervision (2001) The New Capital Accord, Consultative Document, January 2001, BIS.

[25] Bates, J.M. and C.W.J. Granger (1969) The combination of forecasts, *Operational Research Quarterly*, 20, 451-468.

[26] Battiti, R.and A.M. Colla (1994) Democracy in neural nets: voting schemes for classification. *Neural Networks*, 7, 691-707.

[27] Baum, C.F. (2001) A Stata Module to Interpolate a Quarterly Flow Series from Annual Totals via the Proportional DENTON Method. *Mimeo*, Boston college.

[28] Bekaert, A., C.R Harvey and C. Lundblad (2002) Does financial liberalization spur growth? *Working Paper*, Duke University, Durham.

[29] Belkin, B., S. Suchower and L. Forest (1998) A one parameter representation of credit risk and transition matrices. *CreditMetrics Monitor*, 3rd Quarter.

[30] Berg, A. and R. N. Coke (2004) Autocorrelation-corrected standard errors in panel probits: An application to currency crisis prediction. *IMF Working Paper 04*/39.

[31] Berg, A. and J. Sachs (1988) The debt crisis: structural explanations of country performance. *Journal of Development Economics*, 29, 271-306.

[32] Berg, A., and C. Pattillo (1999) Predicting currency crises: The indicators approach and an alternative. *Journal of International Money and Finance*, 18, 561-586.

[33] Bernard, H. and S. Gerlach (1996) Does the term structure predict recessions? The international evidence. *Bank for International Settlements, Working Paper* 37.

[34] Bhargava, A., L. Franzini and W. Narendranathan (1982) Serial correlation and fixed effects model.*Review of Economic Studies*, 49, 533-549.

[35] Bhat, C. (1999) Quasi-random maximum simulated likelihood estimation of the fixed multinomial logit model. *Mimeo*, University of Texas, Austin.

[36] Bhatia, A.V. (2002) Sovereign credit ratings methodology: an evaluation, *IMF Working Paper* 02/170.

[37] Bielecki, B. and M. Rutkowski (2000) Multiple ratings model for defaultable term structure. *Mathematical Finance*, 10, 125-139.

[38] Blume, M.E., F. Lim and A.C. Mackinlay (1998) The declining credit quality of U.S. corporate debt: myth or reality? *The Journal of Finance*, Vol. LIII, 1389-1413.

[39] Breiman, L., J.H. Friedman, E.A. Olsen, and C.J. Stone (1984) *Classification and regression trees,* Belmont, CA: Wadsworth International Group.

[40] Brewer, T.L. and P. Rivoli (1997) Political instability and country risk. *Global Finance Journal*, 8, 309-321.

[41] Bulow, J. and K. Rogoff(1989) Sovereign debt: Is not to forgive to forget? *American Economic Review*, 79, 43-51.

[42] Bunn, D.W. (1975) A Bayesian approach to the linear combination of forecasts, *Operational Research Quarterly*, 26, 325-329.

[43] Burkart, O. and V.Coudert (2002) Leading Indicators of currency crises for emerging countries. *Emerging Markets Review*, 3, 107-133.

[44] Bussiere, M. and M. Fratzscher (2002) Towards a new early warning system of financial crises. *European Central Bank Working Paper 02/1*45.

[45] Callier, P. (1985) Further results on countries' debt-servicing performance: The relevance of structural factors. *Weltwirtschaftliches Archiv*, 121, 105-115.

[46] Cantor, R., and F. Packer (1996) Determinants and Impact of Sovereign Credit Ratings, *FRBNY Economic Policy Review*, October, 37-52.

[47] Carty, L.V. (1993) Moody's rating migration and credit quality correlation, 1920-1996. Moody's Special Report, Moody's Investors' Service, New York.

[48] Carty, L.V. and J.S. Fons (1993) Measuring changes in corporate credit quality. Moody's Special Report, Moody's Investors' Service, New York.

[49] Chamberlain, G. (1980) Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.

[50] Christensen, J., E. Hansen and D. Lando (2004) Confidence sets for continuous-time rating transition probabilities. *Working Paper,* University of Copenhagen.

[51] Claessens, S. and G. Embrechts (2002) Basel II, Sovereign ratings and transfer risk external versus internal ratings. Basel II: An Economic Assessment, Bank for International Settlements, Basel, Conference Proceedings.

[52] Clark, E. and K. Kassimatis (2004) Country financial risk and stock market performance: the case of Latin America. *Journal of Economics and Business*, 56, 21-41.

[53] Clemen, R. R., R. Winkler, and A. Murphy (1995) Screening probability forecasts: contrasts between choosing and combining, *International Journal of Forecasting,* 11, 133-146.

[54] Clemen, R. R. (1989) Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting,* 5, 559-583.

[55] Clemen, R. R. and R. Winkler (1986) Combining economic forecasts, *Journal of Business and Economic Statistics,* 4, 39-46.

[56] Cline, W. (1995) International debt reexamined, Institute of International Economics, Washington.

[57] Cohen, D. (1991) Private lending to sovereign states: A theoretical autopsy, MIT Press, Cambridge, MA.

[58] Cramer, J.S. (1999) Predictive performance of binary logit models in unbalanced samples. *The Statistician,* 48, 85-94.

[59] Crouchy, M., D. Galai and R. Mark (2000) A comparative analysis of current credit risk models. *Journal of Banking and Finance,* 24, 59-117.

[60] Demirguc-Kunt, A. and E.Detragiache (1999) Monitoring banking sector fragility: A multivariate logit approach. *IMF Working Paper* 99/147.

[61] Dempster, A. P., N. M. Laird and D. B. Rubin (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society,* 39, 1-38.

[62] Detragiache, E. and A. Spilimbergo (2001) Crises and liquidity: Evidence and interpretation. *IMF Working Paper* 01/2.

[63] Diebold, F. X and G. D. Rudebusch (1989) Scoring the leading indicators. *Journal of Business,* 62, 369-391.

[64] Diebold, F. X and J. A. López (1996) Forecast evaluation and combination. In *Statistical Methods of Finance*, Handbook of Statistics Series, 14, edited by Maddala, G. S., and C. R. Rao, Elsevier, North-Holland, Amsterdam, New York, pp. 241-268.

[65] Diebold, F. X and P. Pauly (1987) Structural change and the combination of forecasts. *Journal of Forecasting*, 6, 21-40.

[66] Diebold, F. X and R. S. Mariano (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 134-144.

[67] Donkers, B. and B. Melenberg (2002) Testing predictive performance of binary choice models. *Econometric Institute Working Paper.*

[68] Du, Y. (2003) Predicting credit rating and credit rating changes: a new approach. *Working Paper*, Queen's University, Canada .

[69] Eaton, J. and M. Gersowitz (1981) Debt with potential repudiation: Theoretical and empirical analysis. *Review of Economic Studies*, 48, 289-309.

[70] Eaton, J., M.Gersowitz and J.E. Stiglitz (1986) The pure theory of country risk. *European Economic Review*, 30, 481-513.

[71] Edwards, S. (1984) LDC foreign borrowing and default risk: An empirical investigation. *American Economic Review*, 74, 726-734.

[72] Eichengreen, B. and A. Mody (1998) What explains changing spreads on emerging amrket debt: fundamentals or markets entiment? *NBER Working Paper* 6408, Cambridge, Massachusetts.

[73] Elmore, C. and G. McKenzie (1992) Predicting LDC debt arrears. *Mimeo*, University of Southampton, .

[74] Estrella, A. and A. P. Rodrigues (1998) Consistent covariance matrix estimation in probit models with autocorrelated errors. *Mimeo*, Federal Reserve Bank of New York.

[75] Estrella, A. and F. S. Mishkin (1998) Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics* 80, 45-61.

[76] Fair, R. C.and R.J. Shiller (1990) Comparing information in forecasts from econometric models, *American Economic Review*, 80, 375-89.

[77] Feather, P. M. and M. S. Kaylen (1989) Conditional qualitative forecasting. *American Journal of Agricultural Economics*, 71, 195-201.

[78] Feder, G. R., R. E. Just and K. Ross (1981) Projecting debt-servicing capacity of developing countries, *Journal of Financial and Quantitative Analysis* 16, 651-669.

[79] Feder, G.and L. V. Uy (1985) The determinants of international creditworthiness and their policy implications. *Journal of Policy Modelling*, 7, 133-156.

[80] Feder, G., R. E. Just and K. Ross (1981) Projecting debt-servicing capacity of developing countries. *Journal of Financial and Quantitative Analysis*, 16, 651-669.

[81] FitzGerald, V. and D. Krolzig (2003) Modelling the demand for emerging market assets. *Mimeo*, Oxford University, .

[82] Frank, C. R. and W. R.Cline (1971) Measurement of debt-servicing capacity of developing countries: An application of discriminant analysis. *Journal of International Economics*, 41, 327-344.

[83] Frankel, J. A. and A. K. Rose (1996) Currency crashes in emerging markets: An empirical treatment. *Journal of International Economics*, 41, 351-366.

[84] Frydman, H. and A. Kadam (2002) Estimation in the continuous time mover-stayer model with an application to bond ratings migration. *Working Paper,* NYU Stern School of Business.

[85] Glennerster, R. (2004) Transparency and standards: evaluating the effect of institutions, Unpublished PhD thesis, University of London.

[86] Gordy, M. B. (2000) A comparative anatomy of credit risk models. *Journal of Banking and Finance*, 24, 119–149.

[87] Gourieroux, C., A. Monfort and A. Trognon (1984) Estimation and test in probit models with serial correlation. In *Alternative Approaches to Time Series Analysis.* Florens, J. P., Mouchart, M., Raoult, J.P. and Simar, L. (Eds.)

[88] Granger, C. W. J. and M. H. Pesaran (2000) Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19, 537-560.

[89] Granger, C. W. J. and R. Ramanathan (1984) Improved methods of forecasting. *Journal of Forecasting*, 3, 197-204.

[90] Greene, W. H. (2003). *Econometric analysis*, 5th ed., New Jersey: Prentice Hall.

[91] Gupta, S. and P. Wilton (1988) Combination of economic forecasts: an odds-matrix approach, *Journal of Business and Economic Statistics*, 6, 373-379.

[92] Gweke, J., R. C. Marshall and G. A. Zarkin (1986) Mobility indices in continuous-time Markov chains. *Econometrica*, 54, 1407-1423.

[93] Hajivassiliou, V. (1987) The external debt repayment problems of LDS's: An econometric model based on panel data. *Journal of Econometrics*, 36, 205-230.

[94] Hajivassiliou, V. (1989) Do the secondary markets believe in life after debt? In Dealing with Debt Crisis, edited by I. Diwan and I. Hussain.Washington: The World Bank.

[95] Hajivassiliou, V. (1994) A simulation estimation analysis of the external debt crises of developing countries. *Journal of Applied Econometrics*, 9, 109-113.

[96] Hansen, L. S. (1982) Large sample properties of Generalized Method of Moments estimators. *Econometrica*, 50, 1029-1054.

[97] Haque, N., U. Mathieson and N. Mark (1996) The economic content of indicators of developing country creditworthiness. *IMF Staff Papers 96*/43.

[98] Helwege, J. and P. Kleiman (1997) Understanding aggregate default rates on high-yield bonds. *Journal of Fixed Income*, 7, 55-61.

[99] Hsiao, C. and B. H. Sun (2000) To Pool or Not to Pool Panel Data. In Panel Data Econometrics: Future Directions, Papers in Honor of Professor Pietro Balestra, edited by J. Krishnaknmar and E. Rochetti. Amsterdam: North-Holland.

[100] Hu, Y-T., R. Kiesel and W. Perraudin (2002) The estimation of transition matrices for sovereign credit ratings. *Journal of Banking and Finance*, 26, 1383–1406.

[101] Israel, R. B., S. Rosenthal and J. Z. Wei (2001) Finding generators of Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11, 245-265.

[102] Jackson, P. and W. Perraudin (2000) Regulatory implications pf credit risk modelling. *Journal of Banking and Finance*, 24, 1-14.

[103] Jafry, Y. and T. Schuermann (2004) Measurement, estimation and comparison of credit migration matrices. *Wharton Financial Institutions Center Working Paper.*

[104] Jarrow, R. A., D. Lando and S. M. Turnbull (1997) A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10, 481-523.

[105] JP Morgan (1997) Creditmetrics — Technical Document. JP Morgan Discussion Paper, New York.

[106] Juttner, J. D. and J. McCarthy (1998) Modelling a rating crisis. *Mimeo,* Macquarie University Sidney.

[107] Kalbfleisch, J. D. and R. L. Prentice (2002) *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.

[108] Kamin, S. and K. von Kleist (1999) The evolution and determinants of emerging market credit spreads in the 1990s. *Bank for International Settlements Working Paper* 68, Basel.

[109] Kaminsky, G. and C. M. Reinhart (1999) The twin crises: the causes of banking and balance of payment problems. *American Economic Review*, 3, 473-500.

[110] Kaminsky, G. and C.M. Reinhart (2000) On crises, contagion and confusion. *Journal of International Economics*, 51, 145-168.

[111] Kamstra, M. and P. Kennedy (1998) Combining qualitative forecasts using logit. *International Journal of Forecasting*, 14, 83-93.

[112] Kamstra, M., P. Kennedy and T-K. Suan (2001) Combining bond rating forecasts using logit. *The Financial Review*, 37, 75-96.

[113] Kavvathas, D. (2001) Estimating credit transition probabilities for corporate bonds. *Working Paper*, University of Chicago .

[114] Kennedy, P. (2003) *A Guide to Econometrics.* Blackwell: Oxford.

[115] Kijima, M. and K. Komoribayashi (1998) A Markov chain model for valuing credit risk derivatives. *The Journal of Derivatives*, Fall, 97-108.

[116] Kim, J. (1999) Conditioning the transition matrix. *Risk*, October, 37-40.

[117] Kodde, D. A. and F. C. Palm (1986) Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 50, 1243-1248.

[118] Koyluoglu, H. U. and A. Hickman (1998) Credit risk: reconcilable differences. *Risk*, 11, 56-62.

[119] Kumar, M. S., U.Moorthy and W.Perraudin (2003) Predicting emerging market currency crashes. *Journal of Empirical Finance*, 10, 427-454.

[120] Lando D. (2000) Some elements of rating-based credit risk modelling. In: Jegadeesh, N., B. Tuckman (Eds.) Advanced Fixed Income Valuation Tools, 193-215, John Wiley & Sons, New York.

[121] Lando, D. and T. Skodeberg (2002) Analyzing ratings transitions and rating drift with continuous observations. *Journal of Banking and Finance*, 26, 423-444.

[122] Larrain, G., H. Reisen and J. Von Maltzan (1997) Emerging Market Risk and Sovereign Credit Ratings. OECD Development Centre Technical Paper, 124.

[123] Lee, S. H. (1991) Ability and willingness to service debt as explanation for commercial and official rescheduling cases. *Journal of Banking and Finance*, 15, 5-27.

[124] Lee, S.H. (1993). Are the credit ratings assigned by bankers based on the willingness of the LDC borrowers to repay? *Journal of Development Economics*, 40, 349-359.

[125] Leitch, G. and J. E. Tanner (1991) Econometric forecast evaluation: profits versus the conventional error measures. *American Economic Review*, 81, 580-590.

[126] Li, C. A. (1992) Debt arrears in Latin America: Do political variables matter? *Journal of Development Studies*, 28 (4), 668-688.

[127] López, J. A. (2001) Evaluating the predictive accuracy of volatility models. *Journal of Forecasting*, 20, 87-109.

[128] Lucas, D. and J. Lonski (1992) Changes in corporate credit quality 1970-1990. *The Journal of Fixed Income*, 7-14.

[129] Maddala, G. S., V. K. Srivastava and H. Li (1994) Shrinkage estimators for the estimation of long-run parameters from panel data models. *Working Paper*, Ohio State University .

[130] Manasse, P., N. Roubini, and A. Schimmelpfennig (2003) Predicting sovereign debt crises. *IMF Working Paper* 03/221.

[131] Marashaden, O. (1997) A logit model to predict debt rescheduling by less developed countries, *Asian Economices*, 26, 25-34.

[132] Marrison, C. (2002) *The Fundamentals of Risk Measurement.* McGraw and Hill, New York.

[133] Mascharenhas, B. and O. C. Sand (1985) Combination of forecasts in the international context: predicting debt reschedulings. *Journal of International Business Studies*, 20, 539-552.

[134] McCracken, M. W. (2000) Robust out-of-sample inference, *Journal of Econometrics*, 99, 195-223.

[135] McFadden, D., R. Eckaus, G. Feder, V. Hajivassiliou and S. O'Connell (1985) Is there life after debt? An econometric analysis of the creditworthiness of developing

countries. In International Debt and the Developing Countries, 179-209, edited by Gordon Smith and John Cuddington, Washington: The World Bank.

[136] Min, H. (1998) Determinants of emerging market bond spreads: Do economic fundamentals matter? *World Bank Working Paper.*

[137] Monfort, B. and C. Mulder (2000) Using credit ratings for capital requirements on lending to emerging market economies: possible imapct of a New Basel Accord. *IMF Working Paper 02/59.*

[138] Montgomery, A.L., V. Zarnowitz, R.S. Tsay and G.C. Tiao (1998). Forecasting the US unemployment rate, *Journal of the American Statistical Association*, 93, 478-93.

[139] Moody's Investors Service (2003) Sovereign bond defaults, rating transitions and recoveries (1985-2002). *Special Comment* February.

[140] Mulder, C., R. Perilli and M. Rocha (2002) The role of corporate, legal and macroeconomic balancesheet indicators in crisis detection and prevention. *IMF Working Paper* 02/59.

[141] Newbold, P. and D.I. Harvey (2004). Forecast combination and encompassing. In: *A Companion to Economic Forecasting,* edited by M.P. Clements and D.F. Hendry, Blackwell: UK, pp. 268-83.

[142] Newey, W. K. and K. D. West (1987) A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.

[143] Nickell, P., W. Perraudin and S. Varotto (2000) Stability of rating transitions, *Journal of Banking and Finance*, 24, 203-227.

[144] Norris, J. R. (1997) *Markov Chains*, Cambridge University Press.

[145] Obstfeld, M. and K. Rogoff (1996) Imperfections in international capital markets (Ch.6). In Foundations of International Macroeconomics, Cambridge (MA) and London: MIT Press, 349-428.

[146] Oka, C. (2003) Anticipating arrears to the IMF: Early earning systems. *IMF Working Paper* 03/18.

[147] Oral, M., O. Kettani, J. C. Cosset and M. Daouas (1992) An estimation model for country risk rating. *International Journal of Forecasting*, 8, 583-593.

[148] Pesaran, M. H. and A. Timmermann (1992) A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, 10, 461-465.

[149] Pesaran, M. H. and R. P. Smith (1995) Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68, 79-113.

[150] Pesaran, M. H. and S. Skouras (2002) Decision-based methods for forecast evaluation. In *A companion in economic forecasting* edited by M.P. Clements and D.F. Hendry, Blackwell, 241-267.

[151] Peter, M. (2002) Estimating default probabilities of emerging market sovereigns: A new look at a not-so-new literature. *HEI Working Paper* 02/6, Graduate Institute of International Studies, Geneva.

[152] Poirier, D. J. and P. A. Ruud (1988) Probit with dependent observations. *Review of Economic Studies* 55, 593-614.

[153] Reinhart, C. M. (2001) Default, currency crises and sovereign credit ratings. *World Bank Economic Review*, 16, 151-70.

[154] Rojas-Suarez, L. (2001) Rating banks in emerging markets, *Institute for International Economics Working Paper* 01/6.

[155] Saini, K. and P. Bates (1984) A survey of the quantitative approaches to country risk analysis. *Journal of Banking and Finance*, 8, 341-356.

[156] Saunders, A. and L. Allen (2002) *Credit Risk Measurement*, John Wiley & Sons, New York.

[157] Saunders, A. and M. M. Cornett (2003) *Financial institutions management: A risk management approach*, McGraw-Hill, New York.

[158] Schmidt, R. (1985) Early warning of debt rescheduling. *Journal of Banking and Finance*, 8, 357-370.

[159] Schnatz, B. (1999) The sudden Freeze of the Asian Miracle: The Role of Macroeconomic Fundamentals. *Asia Pacific Journal of Finance*, 2, 1-19.

[160] Shleifer, A. (2003) Will the sovereign debt market survive? Harvard Institure of Economic Research DP 2000.

[161] Shorrocks, A. F (1978) The measurement of mobility. *Econometrica*, 46, 1013-1024.

[162] Shumway, T. (2001) Forecasting bankruptcy more accurately: a simpla hazard model. *The Journal of Business*, 74, 101-124.

[163] Solberg, R. L. (1988) Sovereign rescheduling: *Risk and portfolio management.* London, Unwin Hyman.

[164] Sommerville, R. A. and R. J. Taffler (1995) Banker judgement versus formal forecasting models: The case of country risk assessment. *Journal of Banking and Finance*, 19, 281-297.

[165] Staikouras, S.K. (2005) A chronicle of the banking and currency crises. *Applied Economics Letters*.

[166] Staikouras, S. K. (2005) Multinational banks, credit risk and financial crises: A qualitative response analysis. Emerging Markets Finance and Trade.

[167] Standard & Poor (2001) Sovereign defaults decline through third-quarter 2000. In *Ratings Performance 2000, default, transition, recovery, and spreads* edited by S&P, New York, pp. 93-96.

[168] Stock, J.H., & Watson, M. (2001). A comparison of linear and non-linear univariate models for forecasting macroeconomic time series. In: *Festschrift in honour of Clive Granger,* edited by Engle, R.F. and H. White, pp. 1-44.

[169] Sy, A. N. R (2003) Rating the rating agencies: anticipating currency crises or debt crises? *IMF Working Paper 03/122.*

[170] Taffler, R. J and B. Abassi (1984) Country risk: A model for predicting debt-servicing problems in developing countries. *Journal of the Royal Statistical Society,* Series A, 147, 541-568.

[171] Valdes, R. (1997) Emerging market contagion: evidence and theory. *Working Paper Series* 7, Central Bank of Chile, Santiago.

[172] Wei, J. Z. (2003) A multi-factor, credit migration model for sovereign and corporate debts. *Journal of International Money and Finance*, 22, 709-735.

[173] Winkler, R.L. and Makridakis, S. (1983). The combination of forecasts, *Journal of the Royal Statistical Society, Series A*, 146, 150-57.

[174] Wooldridge, J. M. (1994) Estimation and Inference for Dependent Processes. In: *Handbook of Econometrics,* edited by R.F. Engle and D.L. McFadden, Elsevier North-Holland, Amsterdam, London, New York.

[175] Yatchew, A. and Z. Griliches (1984) Specification error in probit models. *Review of Economics and Statistics*, 66, 134-139.

[176] Zou, H. and Yang Y. (2004) Combining time series models, *International Journal of Forecasting*, 20, 69-84.

Appendix 2.1: Country Composition and Indicators

A1. Emerging and developing economies.

Region (number of countries)	Composition
Eastern Europe (7)	Bulgaria, Czech Republic (R), Hungary, Poland, Romania, Russia, Turkey.
South/East Asia (17)	Bangladesh, China, Fiji, India, Indonesia, Korea R, Maldives, Nepal, Pakistan, Papua New Guinea, Philippines, Samoa, Solomon Islands, Sri Lanka, Thailand, Vanuatu, Vietnam.
Latin America (26)	Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican R, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, St. Kitts-Nevis, St. Lucia, Trinidad-Tobago, Uruguay, Venezuela.
Middle East/North Africa (10)	Algeria, Egypt, Iran, Jordan, Lebanon, Morocco, Oman, Syria, Tunisia, Yemen.
Africa (36)	Benin, Botswana, Burkina-Faso, Burundi, Cameroon, Cape Verde, Centr Afr R, Chad, Congo DR, Congo R, Cote d'Ivoire, Eq Guinea, Gabon, Gambia, Ghana, Guinea, Kenya, Lesotho, Malawi, Mali, Mauritania, Mauritius, Mozambique, Niger, Nigeria, Rwanda, Sao Tome Principe, Senegal, Seychelles, Sierra Leone, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe.

A2. Macroeconomic and financial indicators.

Country-specific fundamentals				*Global factors*
External credit exposure	External econ. activity	Domestic conditions	Global links	
Debt/GDP	Export growth[a]	Credit to private sector/GDP	Trade[g]/GDP	World interest rates[j]
Official debt/Total debt	Vol. of export growth[b]	GDP growth[a]	Net bond flow[f,h]	OECD GDP growth[k]
Short-term debt/Reserves[d]	Trade balance[c]/GDP	GNP per capita (1995=100)	Net equity flow[f,h]	US macro. uncertainty
Short-term debt/Total debt	Reserves growth[a,d]	Volatility of GNP pc growth[b]	FDI/GDP[i]	US monet. policy unc.
Debt service/Exports	Reserves/Imports[d]	Gov. expenditure[e]/GDP		US risk aversion index
IMF credit/Exports		Inflation		
		M2/Reserves[d]		
		Real exch. rate (1995=100)[f]		
		Gross capital formation/GDP		
		Gross domestic savings/GDP		

[a] Annual percentage growth. [b] Volatility proxied by the standard deviation over the last four years. [c] Trade balance is total exports - imports. [d] Foreign exchange reserves, excl. gold. [e] Government exp. on consumption, national security and defence. [f] Deviation from long-run trend ($\tilde{x}_{it} = x_{it} - \bar{x}_{i,t-1}$), undervaluation if $\tilde{x}_{it} > 0$. [g] Trade = exports + imports. [h] US$ billion. [i] Foreign direct investment. [j] GDP weighted lending rate for G7 [k] Growth of real GDP per capita for high-income OECD members (GNP p.c. in 1999 $\geqslant$ $9, 361).

Appendix 2.2: Historical Sovereign Defaults per Country 1984-2002

Country	Entries to Default (Δd_{it}=1)	Average Length	Defaults (d_{it}=1)	Default Episodes
Algeria	1	5.0	5	1994-1998
Argentina	4	2.3	9	1984, 1986, 1988-1992,1994-1995
Bangladesh	0	0.0	0	—
Belize	1	1.0	1	1984
Benin	3	2.3	7	1984-1988,1991,1993
Bolivia	4	2.3	9	1984-1985, 1987, 1991-1993, 1995-1997
Botswana	0	0.0	0	—
Brazil	3	1.7	5	1987, 1989-1991, 1993
Bulgaria	2	2.5	5	1990-1993, 1997
Burkina Faso	3	2.7	8	1986-1987, 1992-1994, 2000-2002
Burundi	0	0.0	0	—
Cameroon	3	4.3	13	1986-1988, 1990-1996, 1998-2000
Cape Verde	3	1.7	5	1989-1990, 1993, 1999-2000
Centr Africa R	5	1.6	8	1989-1990, 1992-1993,1995,1998,2000-2001
Chad	2	3.0	6	1985-1987, 1996-1998
Chile	0	0.0	0	—
China	0	0.0	0	—
Colombia	1	1.0	1	1988
Congo DR	2	4.5	9	1988-1995, 1998
Congo R	3	5.3	16	1985, 1987-1993, 1995-2002
Costa Rica	3	2.0	6	1986-1989, 1991, 1993
Cote D'Ivoire	3	3.7	11	1988-1993, 1995, 1998-2001
Czech Rep	0	0.0	0	—
Dominican R	5	2.2	11	1984-1985, 1987-1990, 1992-1993,1995-1996, 1998
Ecuador	3	3.7	11	1987-1994, 1999, 2001-2002
Egypt	3	4.3	13	1984-1986, 1988, 1992-2000
El Salvador	2	2.0	4	1984, 1989-1991
Eq Guinea	4	2.8	11	1984, 1986-1992, 1994-1995, 1998
Fiji	0	0.0	0	—
Gabon	4	2.8	11	1986, 1989-1993, 1995-1998, 2000
Gambia	1	2.0	2	1984-1985
Ghana	1	2.0	2	2001-2002
Grenada	1	8.0	8	1984-1991
Guatemala	3	1.7	5	1986-1987, 1990-1991, 1994
Guinea	4	2.8	11	1985, 1988, 1990-1994, 1996-1999
Guyana	3	3.3	10	1984-1989, 1994-1996, 1999
Haiti	2	2.0	4	1992-1994, 1996
Honduras	5	2.4	12	1984-1986, 1989, 1992-1994, 1996-1997, 1999-2001
Hungary	0	0.0	0	—
India	0	0.0	0	—
Indonesia	1	4.0	4	1998-2001

(cont.)

Iran	1	3.0	3	1984-1986
Jamaica	2	3.5	7	1986, 1989-1993, 1995
Jordan	2	6.0	12	1989-1992, 1994-2001
Kenya	2	2.0	4	1992-1993,2000-2001
Korea	0	0.0	0	—
Lebanon	1	3.0	3	1988-1990
Lesotho	0	0.0	0	—
Malawi	1	1.0	1	1989
Maldives	0	0.0	0	—
Mali	3	3.3	10	1984, 1989-1992,1994-1998
Mauritania	3	4	12	1984, 1989-1995, 1997-2000
Mauritius	0	0.0	0	—
Mexico	1	4.0	4	1989-1992
Morocco	3	2.0	6	1985, 1987, 1989-1992
Mozanbique	3	5.0	15	1984-1986, 1988-1998, 2000
Nepal	0	0.0	0	—
Nicaragua	2	6.5	13	1985-1994, 1997-1999
Niger	4	2.3	9	1989-1990, 1992-1993, 1995, 1997-2000
Nigeria	3	4.0	12	1988, 1990-1999, 2001
Oman	0	0.0	0	—
Pakistan	1	4.0	4	1998-2001
Panama	2	4.0	8	1987-1991, 1993-1995
Papua New Guinea	0	0.0	0	—
Paraguay	2	2.0	4	1986-1987, 1989-1990
Peru	3	3.7	11	1984-1990, 1993-1995, 1998
Philippines	1	5.0	5	1989-1993
Poland	3	2.7	8	1984, 1986-1991, 1997
Romania	0	0.0	0	—
Russia	2	5.5	11	1990, 1992-2001
Rwanda	2	2.5	5	1994-1995, 1999-2001
Samoa	0	0.0	0	—
Sao Tome and Principe	3	4.3	13	1985-1993, 1997-1999, 2001
Senegal	2	3.5	7	1989-1994, 1997
Seychelles	2	1.0	2	1991, 2001
Sierra Leone	4	3.0	12	1985, 1987-1991, 1993, 1996-2000
Solomon Islands	1	9.0	9	1993-2001
Sri Lanka	1	1.0	1	1996
St. Kitts and Nevis	1	1.0	1	1992
St. Lucia	0	0.0	0	—
Swaziland	0	0.0	0	—
Syria	2	6.5	13	1986,1990-2001
Tanzania	4	3.3	13	1984-1985,1987,1989-1996,1998-1999
Thailand	0	0.0	0	—
Togo	4	2.8	11	1987,1989-1994,1996,1998-2000
Trinidad and Tobago	1	5.0	5	1988-1992

(cont.)

Tunisia	0	0.0	0	—
Turkey	0	0.0	0	—
Uganda	3	2.7	8	1988-1992, 1998, 2000-2001
Uruguay	0	0.0	0	—
Vanuatu	0	0.0	0	—
Venezuela	1	2.0	2	1984-1985
Vietnam	2	5.5	11	1988-1996, 1998-1999
Yemen	3	3.7	11	1987-1992, 1995, 1998-2001
Zambia	5	2.4	12	1985, 1987-1990, 1992-1993, 1996-1998, 2000-2001
Zimbabwe	1	3.0	3	2000-2002
Total	175		539	
1984-1995	127		383	
1996-2002	48		156	
Rate	10%		30%	
1984-1995	11%		33%	
1996-2002	7%		23%	

The models have the form $y_{it}=f(x_{i,t-1})$ so the first relevant year for y_{it} in the analysis is 1984. The reported statistics are for the default series $\{d_{it}\}_{t=1984}^{2002}$ on which the EWS indicator $\{y_{it}\}_{t=1984}^{2000}$ is based, e.g. $y_{i,2000}=1$ if $d_{i,t}=1$ at t=2000,2001 or 2002. A country-period (i,t) case is a 'default entry' if $d_{i,t-1}=0$ and $d_{it}=1$. The reported default entries in 1984 are cases where $d_{i,1983}=0$. The analysis is based on N=96 countries. There are 1152(=96×12) and 672(=96×7) country-period cases over 1984-1995 and 1996-2002, respectively.

Appendix 2.3: The Cross-Validation (Jacknife) Method

In order to preserve degrees of freedom, a jacknife procedure is conducted to reduce the original set of explanatory variables to an optimal smaller set with large predictive power. This jacknife approach is conducted in-sample, i.e. over the 1984-1995 period denoted $[1, T^*]$. It is based on the Type I error measured over the entire sample $[1, T^*]$ and over a reduced subset that excludes consecutive defaults. The former measure (denoted TI) gives the percentage of missed *defaults* $(\hat{y}_{T_0+1} = 0, y_{T_0+1} = 1)$ whereas the latter gives the percentage of mispredicted positive directional changes (PDC) or *entries to default* $(\hat{y}_{T_0+1} = 0, y_{T_0+1} = 1, y_{T_0} = 0)$.

The pooled logit model estimates over $[1, T_0]$ with $T_0 < T^*$ and $\lambda = 0.5$ are used to generate 1-step-ahead event forecasts $\hat{y}_{i,T_0+1}$ for $i = 1, .., N$ (minimum feasible $T_0 = 4$). This modeling and forecasting exercise is repeated iteratively, adding one further observation at a time, until T^* is reached. We compute the following cross-validation (CV) metric for different regressor sets (S)

$$CV\_TI_S = \frac{1}{(T^* - T_0)} \sum_{t=T_0+1}^{T^*} TI_t$$

and likewise for CV_PDC_S. In the first iteration, the baseline regressor set S_0 contains all regressors and S_j is a model that differs from S_0 in that it excludes x_j. Each iteration has 2 steps. First, collect in $\tilde{X}$ the $x_j \in S_0$ that satisfy

$$CV\_PDC_{S_j} \leq CV\_PDC_{S_0}$$

so that PDC is not increased by excluding any of them. Second, collect in $\breve{X}$ the $x_k \in \tilde{X}$ such that

$$CV\_TI_{S_k} \leq CV\_TI_{S_0}$$

so that their exclusion does not increase TI. The regressor set S_r that satisfies

$$S_r = \arg\min_{k \in \breve{X}} (CV\_TI_{S_k} - CV\_TI_{S_0})$$

in the first iteration is the reduced regressor set that gives the minimal TI without increasing PDC relative to that for S_0. Therefore, x_r is dropped from S_0 and the new baseline regressor set for the second iteration is S_r and so forth. The last iteration occurs when $\breve{X}$ is the null set. We thus end up with a regressor set that gives the smallest possible Type I error over $[1, T^*]$ under the condition that no variable can be removed without increasing the Type I error over the PDC sample.

Appendix 2.4: Definition of Selected Variables

External credit exposure

External debt/GDP: The total external debt measured as a share of GDP. Total external debt is debt owed to nonresidents repayable in foreign currency, goods, or services. Total external debt is the sum of public, publicly guaranteed, and private nonguaranteed long-term debt, use of IMF credit, and short-term debt.

Official debt/Total debt: Public and publicly guaranteed debt comprises long-term external obligations of public debtors, including the national government, autonomous public bodies, and external obligations of private debtors that are guaranteed for repayment by a public entity.

Short-term debt/Total debt: Short-term debt includes all debt having an original maturity of one year or less and interest in arrears on long-term debt.

Debt service ratio: Total debt service relative to exports. Total debt service is the sum of principal and interest repayments paid in *foreign currency* on long-term debt, interest paid on short-term debt and repayments to the IMF.

IMF credit/Exports: The repurchase obligations to the IMF for all uses of IMF resources as a share of export earnings. These obligations comprise purchases outstanding under the credit tranches, including enlarged access resources, and all special facilities (extended fund, and oil facilities), trust fund loans, and operations under the structural adjustment facilities.

External economic activity

Variability of export growth: The standard deviation of annual growth rate of exports of goods and services over the past 5 years. Annual growth rate of exports of goods and services is based on constant local currency. Aggregates are based on constant 1995 U.S. dollars. Exports of goods and services represent the value of all goods and other market services provided to the rest of the world (e.g. merchandise, insurance, transport, travel, communication, construction, financial, business and government services).

Trade balance /GDP: Trade balance relative to GDP. Trade balance is defined as the difference between exports and imports.

Domestic conditions

Credit to private sector/GDP: Credit to private sector includes the domestic financial resources, provided to the private sector, such as loans, purchases of non-equity securities, trade credits and other accounts receivable that establish a claim for repayment.

GDP growth: The annual growth rate of the gross domestic product. It is based on constant local currency. Aggregates are based on constant 1995 U.S. dollars.

GNP per capita: GNP per capita is the gross national product divided by midyear population. GNP is the sum of value added by all resident producers plus any product taxes not included in

the valuation of output plus net receipts of primary income from abroad. Data are in constant 1995 U.S. dollars.
Variability of GNP per capita growth: The standard deviation of GNP per capita annual growth rate over the past 4 years. Annual growth rate of GNP per capita is based on constant local currency. Aggregates are based on constant 1995 U.S. dollars.
Real exchange rate misalignment: Deviation of real exchange rate from the long-run trend. The variable was constructed using the official exchange rate (local currency per US$) , determined by national authorities, adjusted for relative changes in the price levels. It is calculated as an annual average based on monthly averages. Our measure of the long-run trend is the demeaned real exchange rate index.

Global Links

Trade/GDP: The trade integration ratio is the sum of exports and imports of goods and services measured as a share of GDP converted to international $ using PPP rates.

Global Factors

US monetary policy uncertainty: The measure is the conditional standard deviation derived from an AR(1)-GARCH(2,1) model fitted to the spread between the yield on the U.S Federal Funds Target Rate and the yield on the 3 month US Tresaury Bill.
US macroeconomic uncertainty: The measure of macroeconomic uncertainty is the conditional variance derived from an AR(1)-GARCH (1,1) model fitted to the detrended (logarithmic first differences) U.S Real GDP.
US risk aversion: The measure for risk aversion is the Sharpe ratio of the spread between the Merrill Lynch corporate high yield 175 index and the yield on the 10-year U.S Treasury Bonds.

Appendix 2.5: Expected Impact of Selected Variables

External credit exposure

External debt/GDP (+): An increase in the size of the debt burden, compared to the country's output (or any other type of resources such as export earnings) will increase the likelihood that the debt burden becomes unsustainable and thus, the probability of default (Edwards, 1984; McFadden et al., 1985; Callier, 1985; Hajivassiliou, 1994; Aylward and Thorne, 1998; Detragiaghe and Spilimbergo, 2001)
Official debt/Total debt (+): Detragiaghe and Spilimbergo (2001) argue that the burden of servicing a given amount of debt is likely to depend on the nature of the debt obligations. Thus, one should control for debt characteristics such as the share of debt owed to multilateral creditors, commercial creditors, and the share of debt at concessional terms. They find a positive and significant effect of the official debt burden, while variables that indicate the share of debt

owed to commercial banks or that at concessional terms prove insignificant. The latter result reflects that countries experiencing balance of payments problems are more likely borrowers from multilateral institutions.

Short-term debt/Total debt (+): The ratio is a proxy for short-term external liquidity and affects the level of foreign funds that a country requires to raise in a given year. Detragiaghe and Spilimbergo (2001) find a positive and significant effect, that is, the less liquid a country the more likely it is to default on its external debt obligations.

Debt service ratio (+): The ratio links the fixed foreign exchange outflow obligation of debt service to what is generally the major foreign exchange inflow. As noted in Feder et al. (1981) a shortfall of exports will force the government to draw down foreign exchange reserves or cut down imports in order to accommodate debt-service payments and thus, an increase in the ratio would increase the risk of defaulting by the government. However, Frank and Cline (1971) point out that the ratio may be non-informative as a proxy for the country's ability to service foreign currency debt. The debt service ratio is just an indicator of the proportion of foreign exchange earnings which can be used for imports. If export earnings are high relative to import demand, a high debt service ratio can be maintained without increasing the probability of default. Furthermore, a country with a good credit standing may be able to finance a high debt service ratio, at least for a certain amount of time, through a high level of borrowing. A significant and positive effect is reported in Frank and Cline (1971), Feder et al. (1981), Edwards (1984), Berg and Sachs (1988), Brewer and Rivoli (1997), while a negative coefficient is reported in Elmore and McKenzie (1992).

IMF credit/Exports (+/-): Aylward and Thorne (1998) report a positive impact on default risk to external creditors. This can be interpreted by considering the IMF's preferred creditor status, which implies that other creditors are repaid only after IMF obligations are met. The positive effect of this ratio is also in line with that countries experiencing balance of payments problems are more likely borrowers from multilateral institutions (Detragiaghe and Spilimbergo, 2001). On the other hand, certain countries benefit from last minute IMF rescue packages and prevent default (Manasse et al., 2003), which postulates a negative parameter.

External economic activity

Variability of export growth (+/-): Export earnings is the main source of countries' foreign exchange. According to the views advocated within the ability-to pay framework the more the fluctuations in the flows of foreign funds faced by a country the higher the probability for a large current account deficit, and thus of default. On the other hand, Eaton and Gersowitz (1981) assume that a country uses foreign funds to smooth consumption. Under this assumption, the more the uncertainty in a country's export earnings, the more a country relies on foreign fund flow from international creditors to stabilise consumption. Thus, the greater is the incentive the country has to meet debt service obligations to avoid potential penalties imposed by creditors. In this perspective, the expected parameter is negative. In essence, the theoretical literature provides explanations for both positive and negative effects and there is little established empirical evidence on the effect of export shocks. Lee (1991) is the only empirical study that looks at volatility of exports and finds a negative impact.

Trade balance /GDP (-): It is a proxy for the external balance of payments or current account deficit relative to the country's output. Edwards (1984) argues in favour of a negative effect of this variable which indicates that a higher balance of payment deficit will result in a higher probability of default because the same investment is being financed with a higher proportion of foreign savings. Moreover, for countries with large trade balance deficits, any shock disrupting a country's access to international capital markets will compound the problem of servicing maturing debt with that of finding a substitute for the real resource transfer the country relied upon. Haque et al.(1996), among others, argue that an improving current account position reduces the dependence on foreign savings, slows down the increase in the foreign debt burden or even reduces the foreign debt burden and, thus, reduces default risk. Cline (1984), Edwards (1984), Callier (1985), Hajivassiliou (1994) and Peter (2002) report negative coefficients.

Domestic conditions

Credit to private sector/GDP (+/-): For the impact of the ratio of (domestic) credit to private sector to GDP two contradicting theoretical expectations can be considered. The first scenario expects the ratio to be positively correlated to a country's default risk since the higher the indebtedness of the private sector compared to the output of the economy, the higher the likelihood of mass private bankruptcies, in times of financial distress. A banking crisis may induce a country to bail out large financial institutions, and therefore plunge into debt repayment difficulties (Peter, 2002). From a different perspective, a larger private sector may be reflect better dsitribution of assets and in this case the expected parameter is negative (Staikouras, 2005) . Bekaert et al. (2000) argue that the credit to private sector ratio is a proxy for banking development which is linked with increased growth. Real investment and infrastructure developments that lead to economic growth are carried out by the private sector.

GDP growth (-): Haque et al.(1996), Balkan (1992), Feder and Uy (1985) and McFadden et al. (1985) report a significant negative effect of GDP growth on default risk. Feder and Uy (1985) suggest that higher growth would improve the initial creditworthiness of a country through increased investment, however, as this higher growth entails heavier borrowing for investment it may be possible to reduce creditworthiness in subsequent periods.

GNP per capita (-): The ratio is indicative of a government's management flexibility (Feder et al., 1981). It serves as a proxy of the extent to which the government can muster additional resources to overcome a balance of payment liquidity crisis without defaulting. The higher the per capita income the easier it would be for the government to dampen the demand for goods and services and divert the resources to produce the additional foreign exchange needed to service debt obligations (management flexibility). Feder et al. (1981), Feder and Uy (1985), Berg and Sachs (1988), McFadden et al. (1985), Hajivassiliou (1987), Aylward and Thorne (1998) report significant negative effects.

Variability of GNP per capita growth (+/-): The expectations for the effect of the volatility of GNP per capita growth are overwhelmed by ongoing controversy between the willingness and capacity to pay theories. The arguments of the two sides are similar to the ones discussed for export volatility shocks. According to the Eaton-Gersowitz willingness to pay approach and the consumption-smoothing motive, the expected effect is negative (Lee, 1991). If the capacity to pay considerations prevail the parameter will bear a positive sign since the more the shocks on a country's income the higher the probability that the country might experience output ranges that diminish the resource availability for debt repayment (Peter, 2002). In essence, whether the income volatility is positively or negatively correlated to sovereign default risk depends on the conceptualised trade-off between a country's willingness to pay and capacity-to -pay benefits, and on which of the two will prevail.

Real exchange rate misalignment (+/-): Three of the studies that tested a real exchange rate indicator agree on a significant positive effect on default probability (Solberg, 1988; Obedokun, 1995; Peter, 2002). Other studies (Haque et al., 1996; Hajivassiliou, 1989; Detragiaghe and Spilimbergo, 2001) report a positive but insignificant effect. Apreciation of the currency is likely to lead to currency crises, while depreciation will imply large external debt burden in terms of domestic currency (Peter, 2002). Detragiaghe and Spilimbergo (2001) refer to the negative effect that an overvalued exchange rate may have on exports. Likewise, Haque et al. (1996) note that the real exchange rate variable measures the trade competitiveness of the economy, thus a high real exchange rate is expected to increase default risk. On the other hand, Manasee et al. (2003) find that the exchange rate shows a large depreciation against the US dollar in the year of entry into a debt crisis, and argue that depreciation increases external debt in domestic currency units which can make debt repayments more difficult.

Global Links

Trade/GDP (-): It indicates the degree of trade openness of an economy. It has been found significant and negative in Detragiaghe and Spilimbergo (2001) and Callier (1985). Trade integration increases growth through export revenues (Bekaert et al., 2000). Furthermore, un a willingness-to pay framework more trade openness may render a country more vulnerable to creditor sanctions if it defaults (Bulow and Rogoff, 1989).

Global Factors

The empirical literature on international capital flows gives support to the impact of external factors on country creditworthiness and on the fluctuations of capital flows to emerging markets. Business cycle fluctuations in industrialized countries (namely higher interest rates and lower availability of capital) and changes in risk aversion induce shifts in the demand of emerging market assets, which in turn influence capital flows (FitzGerald and Krolzig, 2003). In the

context of debt servicing capacity this translates into lower levels of foreign exchange reserves. Higher interest rates mean higher debt-service for the newly acquired debt of emerging market.
US monetary policy uncertainty (+): Fluctuations in the spread are aimed at capturing heightened uncertainty about the expected stance of U.S monetary policy. Arora and Cerisola (2000) find that US monetary policy uncertainty increases sovereign spreads, since unpredictability of US monetary policy destabilize capital flows and capital market conditions in emerging markets. Some authors have argued that episodes of market turbulence reflect irrational investor behaviour and contagion (Valdes, 1997; Kaminsky and Reinhart, 2000), while other have tried to explain these episodes primarily as liquidity events. Other studies (Edwards and Susmel, 2000) have stressed the importance of the transmission mechanism of such liquidity effects on capital flows and prices of emerging market assets. Sudden and unexpected rises in the US interest rates or other shocks precipitate financial turmoil across assets and countries.
US macroeconomic uncertainty (+): Abrupt business cycle fluctuations in industrialized countries change market sentiment (through spillover of financial turmoil or contagion of shocks from one country to another) and lower the demand of emerging market assets, which in turn reduce capital flows.
US risk aversion (+): Risk aversion varies over time and negatively affects capital flows to emerging markets as lower appetite for risk implies a reduction in the demand and the exposure to emerging market risky assets(FitzGerald and Krolzig, 2003). Changes in market sentiment have been evidenced to drive fluctuations in emerging market spreads (Cantor and Packer; 1996; Eichengreen and Mody, 1998; Kamin and von Kleist, 1999).

Appendix 2.6: Cross-Section and Regional Estimates

Variables	TCS Min	TCS Max	TCS Median	RLOGIT I	RLOGIT II	RLOGIT III	RLOGIT IV	RSLOGIT I	RSLOGIT II	RSLOGIT III	RSLOGIT IV
External debt/GDP (+)	4.24	29.41	10.85	23.03	6.17	7.18	5.20	13.75	7.96	5.86	—
	(2.17)	(3.10)	(3.29)	(3.61)	(3.84)	(8.72)	(2.24)	(5.15)	(8.34)	(9.23)	
Offic debt/ Tot debt (+)	-12.71	18.61	8.73	10.26	15.67	4.19	30.23	-0.88	16.94	8.45	37.78
	(-0.97)	(1.07)	(0.68)	(1.48)	(3.27)	(1.06)	(2.43)	(-0.25)	(6.33)	(2.81)	(3.65)
ST debt/ Tot debt (+)	-14.94	16.25	5.10	7.99	2.83	14.63	17.14	-2.55	—	15.13	24.24
	(-1.33)	(2.01)	(0.79)	(0.99)	(0.74)	(3.92)	(1.80)	(-0.53)		(4.92)	(3.04)
Debt serv/Exports (+)	-24.72	7.32	-5.22	-3.63	-3.09	-5.92	-7.50	—	-3.50	-2.82	—
	(-2.87)	(1.53)	(-0.96)	(-0.57)	(-1.40)	(-3.39)	(-2.72)		(-2.24)	(-2.16)	
IMF credit/Exports (+/-)	-19.26	10.45	-1.65	-6.04	5.36	-2.07	11.63	-4.62	—	-1.52	14.60
	(-2.50)	(2.27)	(-0.42)	(-1.07)	(1.88)	(-1.73)	(1.69)	(-1.58)		(-1.85)	(2.43)
Vol export growth (+/-)	-12.14	12.89	1.34	-13.48	1.78	0.21	4.65	—	—	—	4.37
	(-2.24)	(1.93)	(0.36)	(-2.13)	(0.85)	(0.17)	(1.78)				(1.79)
Trade balance/GDP (-)	-6.59	8.61	0.83	-2.45	1.72	0.03	-0.74	—	—	—	—
	(-1.16)	(1.94)	(0.17)	(-0.36)	(0.73)	(0.03)	(-0.24)				
Credit private/GDP (+/-)	-14.33	-0.86	-4.50	-1.47	-3.71	-0.81	-3.76	—	—	—	-4.73
	(-2.61)	(-0.33)	(-1.30)	(-0.29)	(-2.92)	(-0.45)	(-1.83)				(-3.01)
GDP growth (-)	-18.16	17.84	-7.89	10.64	0.91	-0.48	1.04	—	—	—	—
	(-1.54)	(1.40)	(-1.00)	(1.18)	(0.25)	(-0.21)	(0.23)				
GNP per capita (-)	-0.36	2.74	0.98	1.12	-0.33	0.63	-1.92	-0.51	—	0.53	-2.93
	(-0.58)	(2.97)	(1.90)	(1.04)	(-0.92)	(2.86)	(-2.47)	(-0.82)		(2.97)	(-4.99)
Vol pc growth (+/-)	-29.21	27.89	-0.73	-0.53	4.91	7.62	-2.64	-3.69	4.53	—	—
	(-1.20)	(1.62)	(-0.05)	(-0.02)	(0.60)	(2.00)	(-0.27)	(-0.29)	(0.84)		
Real exchange rate (+/-)	-3.38	1.53	0.49	0.54	0.55	-0.18	-0.85	—	—	—	—
	(-2.16)	(1.45)	(0.83)	(0.51)	(1.55)	(-0.67)	(-1.21)				
Trade/GDP (-)	-19.37	0.43	-6.54	-12.60	-5.49	-7.19	-3.29	-6.59	-7.46	-6.76	2.74
	(-2.87)	(0.16)	(-2.18)	(-1.79)	(-3.82)	(-6.64)	(-1.21)	(-2.52)	(-7.39)	(-7.57)	(1.79)

t-statistics are reported in parenthesis. (I) Asia, (II) Latin America, (III) Africa, (IV) East Europe/Middle East/North Africa.

Appendix 4.1: Clustering by K-means Algorithm

K-means clustering belongs to the non-hierarchical clustering class of methods. In our context, a case is an observation vector $\mathbf{x}_{it} = [x_{it,1}, x_{it,2}, ..., x_{it,s}]'$ where s is the dimension of the macrovariable set; $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ denote country and period, respectively. The number of sample cases is $M = NT$.

K-means clustering consists of comparing the distances of each observation vector from the mean vectors of each of K proposed clusters in the sample of M observations. The observation $\mathbf{x}_{it}$ is assigned to the cluster with nearest mean vector. The distances are recomputed and reassignments are made as necessary. This process continues until all observations are in clusters with minimum distances to their mean vectors. The algorithm in steps is as follows:

1. Take the first K cases in the sample as the initial cluster centroids ($\mathbf{c}_1^0 \equiv \mathbf{x}_{11}, \mathbf{c}_2^0 \equiv \mathbf{x}_{21}, ..., \mathbf{c}_K^0 \equiv \mathbf{x}_{K1}$).

2. Assign case $\mathbf{x}_{it}$ to the cluster whose centroid is closer

$$\mathbf{c}_j^0 = \underset{q=1,2,...,K}{\operatorname{argmin}} D(\mathbf{x}_{it}, \mathbf{c}_q^0)$$

$i = 1, 2, ..., N, t = 1, 2, ..., T$ where $D(\mathbf{x}_{it}, \mathbf{c}_q^0)$ denotes the Euclidean distance between the it^{th} case and q^{th} cluster centroid, given by

$$D(\mathbf{x}_{it}, \mathbf{c}_q^0) = \sqrt{\sum_{l=1}^{s} (x_{it,l} - c_{q,l}^0)^2}$$

Thus the outcome of Step 2 is a set of K clusters of observation vectors. Let $m_1, m_2, ..., m_K$ denote the number of observation vectors in each cluster such that $\sum_{q=1}^{K} m_q = M$.

3. The centroid of the new qth cluster is given by its mean observation vector

$$\mathbf{c}_q^1 = \left[\ \tfrac{1}{m_q} \textstyle\sum_{it} x_{it,1}, ..., \tfrac{1}{m_q} \textstyle\sum_{it} x_{it,s} \ \right]'$$

which facilitates a measure of the change in the cluster centroids, $\Delta S_q = D(\mathbf{c}_q^1, \mathbf{c}_q^0)$, $q = 1, 2, ..., K$.

4. If $\Delta S_q < \varepsilon$ for all $q = 1, 2, ..., K$ the algorithm terminates. Otherwise it goes to Step 2. We set $\varepsilon = 0.01$. The output of the procedure is the set of K clusters obtained at iteration j such that $\Delta S_q = D(\mathbf{c}_q^{j+1}, \mathbf{c}_q^j) < \varepsilon$ for all $q = 1, 2, ..., K$.

Appendix 4.2: Historical Sovereign Defaults per Year, 1984-2000

Year	Default entries (Δd_{it}=1)	Countries
1984	15 (7)	Argentina, Benin, Bolivia, Dominican R, Egypt, El Salvador, Grenada Honduras, Iran, Mali, Mozanbique, Peru, Poland, Tanzania, Venezuela
1985	6 (3)	Congo R, Guinea, Morocco, Nicaragua, Sierra Leone, Zambia
1986	9 (7)	Argentina, Burkina Faso, Cameroon, Costa Rica, Gabon, Guatemala, Jamaica, Paraguay, Syria
1987	11 (9)	Bolivia, Brazil, Congo, Dominican R, Ecuador, Morocco, Panama, Sierra Leone, Tanzania, Togo, Zambia
1988	12 (7)	Argentina, Colombia, Congo DR, Cote D'Ivoire, Egypt, Guinea, Lebanon, Mozanbique, Nigeria, Trinidad-Tobago, Uganda, Vietnam
1989	15 (10)	Brazil, El Salvador, Gabon, Honduras, Jamaica, Jordan, Malawi, Mali, Mexico, Morocco, Paraguay, Philippines, Senegal, Tanzania, Togo
1990	7 (4)	Bulgaria, Cameroon, Guatemala, Guinea, Nigeria, Russia, Syria
1991	4 (3)	Benin, Bolivia, Costa Rica, Seychelles
1992	8 (5)	Burkina Faso, Dominican R, Egypt, Haiti, Honduras, Kenya, Russia, Zambia
1993	6 (5)	Benin, Brazil, Costa Rica, Panama, Peru, Sierra Leone
1994	5 (5)	Algeria, Argentina, Guatemala, Jordan, Mali
1995	6 (6)	Bolivia, Congo R, Cote D'Ivoire, Dominican R, Gabon, Jamaica,
1996	7 (7)	Guinea, Haiti, Honduras, Sierra Leone, Sri Lanka, Togo, Zambia
1997	4 (4)	Bulgaria, Nicaragua, Poland, Senegal
1998	11 (11)	Cameroon, Congo DR, Cote D'Ivoire, Dominican R, Indonesia, Pakistan, Peru, Tanzania, Togo, Uganda, Vietnam
1999	2 (2)	Ecuador, Honduras
2000	7 (7)	Burkina Faso, Gabon, Kenya, Mozanbique, Uganda, Zambia, Zimbabwe

	Total	Rate
1984-2000	135 (102)	10% (10%)
1984-1995	104 (71)	11% (11%)
1996-2000	31 (31)	8% (8%)

Data on debt levels, arrears and reschedulings for N=75 countries over T=17 years (1984-2000) is used to define NT=75×17=1275 cases for the default indicator d_{it}. Numbers in parentheses pertain to the effective country-period cases ($\widetilde{NT}$=1017) used in the analysis; $\widetilde{NT}$ is dictated by the missing cases for X (=macrovariables or credit ratings). The following countries did not experience default: Bangladesh, Bostwana, Chile, China, Czech R, Ghana, Hungary, India, Korea, Mauritus, Nepal, Oman, Papua Guinea, Romania, Swaziland, Thailand, Tunisia, Turkey, Uruguay.

Appendix 4.3: Logit Estimates 1984-1995

Variable	Classifier LOGIT	LOGIT-II
External debt/ GDP (+)	*6.959* (7.19)	–
Official debt/ Total debt (+)	*2.473* (2.03)	–
Short term debt/ Total debt (+)	-0.068 (-0.06)	–
Debt service/ Exports (+)	-1.940 (-3.13)	–
IMF credit/ Exports (+/-)	*-1.206* (-1.78)	–
Volatility export growth (+/-)	*1.415* (2.18)	–
Trade balance/ GDP (-)	*-1.198* (-1.34)	–
Credit private sector/ DGP (+/-)	*-1.799* (-3.31)	–
GDP growth (-)	*-0.813* (-1.39)	–
Per capita GNP (-)	1.914 (3.31)	–
Volat. pc GNP growth (+/-)	*1.681* (3.08)	–
RER misalignment (+/-)	*0.437* (0.76)	–
Total trade/ GDP (-)	*-2.751* (-3.99)	–
II credit ratings (-)	–	*-0.082* (-9.53)
Intercept	-2.641 (-1.60)	1.406 (6.66)

The model is $y_{it} = f(x_{i,t-1})$, where $y_{it} \equiv d_{it}$.Expected theoretical sign in parenthesis besides each variable; italics indicate correct sign. Volatility defined as std. deviation over [t-3,t]. RER misalignement is the real exchange rate deviation from trend; trend proxied by the sample mean up to t-1. In parenthesis the t-statistics are reported.

Appendix 5.1: Bootstrapping the Ratings Migration Process

The bootstrap simulation method in steps is as follows:

1. Estimate the generator matrix Λ from the observed sample of sovereign ratings (N sovereigns)

2. Calculate the probability transition matrix as $P(\Delta t) = \exp(\Lambda \Delta t)$. Choose the transition horizon Δt for the analysis. Transform the transition matrix into cumulative probability ranges for each rating.

3. The initial rating X_{0i} and and lifetime h_i for each sovereign ($i = 1, ..., N$) are those in the observed sample. Obtain the next period state, X_{1i}, by generating a random draw $r \sim i.i.d.U[0,1]$ and matching it with the cumulative probability ranges corresponding to X_{0i}. For instance, if $r \in (p_{11} + p_{12}, p_{11} + p_{12} + p_{13}]$, then the rating associated with the 3^{rd} column of P represents X_{1i}.

4. Repeat step 3 with starting state X_{1i} to simulate X_{2i} and so forth. The artificial rating history for sovereign i is $\{X_{0i}, X_{1i}, ..., X_{h_i}\}$

5. Repeat steps 3 and 4 for all sovereigns $i = 1, ..., N$. The output is the bootstrap dataset B_j that contains the N rating histories.

6. Transform B_j into sequences of rating transition and durations. Use the latter to compute the transition probability matrix $\hat{P}(\Delta t)$ for the horizon of interest using the DTM and HHR estimation methods. Retain the default probability vector $P\hat{D}_j$ and compute the mobility differential $\Delta \hat{m}_j$.

7. Repeat steps 3 to 6 a large number, R, of times. The parameters of interest are the R default probability vectors $\left\{P\hat{D}_j\right\}_{j=1}^{R}$ and R mobility differential scalars $\{\Delta \hat{m}_j\}_{j=1}^{R}$ from the DTM and HHR estimation methods.

Appendix 5.2: Moody's Foreign-Currency Sovereign Bond Issuers

A. Characteristics of Moody's-rated foreign currency sovereign bond issuers

						Level		Region				
Year	Rated sovereigns	IG	NIG	WR	Default	Industrial	Emerging	Asia	Latin Amer.	East. Europe	Africa	Mid. East
1981	11	11	0	0	0	9	2	0	2	0	0	0
1982	11	11	0	0	0	9	2	0	2	0	0	0
1983	11	9	0	2	0	9	2	0	2	0	0	0
1984	11	9	0	2	0	9	2	0	2	0	0	0
1985	11	9	0	2	0	9	2	0	2	0	0	0
1986	11	8	0	3	0	9	2	0	2	0	0	0
1987	16	12	2	2	0	11	5	1	4	0	0	0
1988	19	15	3	1	0	14	5	1	4	0	0	0
1989	21	17	3	1	0	15	6	2	4	0	0	0
1990	23	19	3	1	0	16	7	3	4	0	0	0
1991	24	19	3	2	0	16	8	3	5	0	0	0
1992	24	19	5	1	0	16	8	3	5	0	0	0
1993	26	19	5	2	0	16	10	3	6	0	0	1
1994	31	21	9	1	0	17	14	5	8	0	0	1
1995	36	24	11	1	0	17	19	6	10	1	1	1
1996	40	28	11	1	0	17	23	6	10	2	2	3
1997	49	31	18	0	0	17	32	7	10	8	2	5
1998	57	32	25	0	0	17	40	6	15	12	2	5
1999	69	40	27	0	2 (2)	17	52	7	20	17	3	5
2000	72	42	26	0	4 (2)	17	55	7	21	17	4	6
2001	72	44	24	0	4 (1)	17	55	7	21	17	4	6
2002	72	44	25	0	3 (2)	17	55	7	21	17	4	6
2003	72	45	24	0	3 (1)	17	55	7	21	17	4	6

Ratings are observed on 5th March every year. The rated sovereigns per year are categorised by credit quality and state of the economy.
IG stands for Investment Grade issuers, NIG for Non-Investment Grade issuers and WR for Withdrawn Ratings. The numbers in parentheses are default entries for the given year. The rated emerging markets are further categorised by regional location.

B. List of Moody's-rated foreign currency sovereign bond issuers

Year	New Ratings	Countries
1981-1986	11	Australia, Austria, Canada, Denmark, Finland, New Zealand, Norway, Panama, Sweden, United Kingdom, Venezuela
1987	5	Argentina, Brazil, Italy, Malaysia, Portugal
1988	3	Belgium, Ireland, Spain
1989	2	China, France
1990	2	Iceland, Thailand
1991	1	Mexico
1992	0	None
1993	2	Trinidad and Tobago, Turkey
1994	5	Colombia, Indonesia, Japan, Philippines, Uruguay
1995	5	Barbados, Bermuda, Greece, Pakistan, South Africa
1996	4	Israel, Jordan, Mauritius, Poland
1997	9	Bulgaria, Croatia,Kazakhstan, Lebanon, Lithuania, Moldova, Oman, Russia, Slovenia
1998	8	Bahamas, Costa Rica, Cyprus, Ecuador, Guatemala, Jamaica, Romania, Ukraine
1999	12	Belize, Bolivia, Czech Republic, Dominican Republic, El Salvador, Egypt, Estonia, Hungary, Korea, Latvia, Peru, Slovakia
2000	3	Chile, Morocco, Qatar
2001	0	None
2002	0	None
2003	0	None
Total	72	

Ratings are observed on 5th March every year.

Appendix 5.3: Moody's Ratings Distribution

A. Number of rated sovereigns. Groups by rating and year

	Rating										
Year	Aaa	Aa	A	Baa	Ba	B	C	Default	Total	IG	NIG
1981	8	3	0	0	0	0	0	0	11	11	0
1982	8	3	0	0	0	0	0	0	11	11	0
1983	6	3	0	0	0	0	0	0	9	9	0
1984	6	3	0	0	0	0	0	0	9	9	0
1985	5	4	0	0	0	0	0	0	9	9	0
1986	6	2	0	0	0	0	0	0	8	8	0
1987	8	2	1	1	2	0	0	0	14	12	2
1988	7	6	1	1	1	2	0	0	18	15	3
1989	8	6	2	1	1	2	0	0	20	17	3
1990	7	7	4	1	1	2	0	0	22	19	3
1991	5	9	4	1	2	1	0	0	22	19	3
1992	4	10	4	1	2	2	0	0	23	19	4
1993	4	10	4	1	3	2	0	0	24	19	5
1994	5	9	6	1	7	2	0	0	30	21	9
1995	4	11	6	3	9	2	0	0	35	24	11
1996	4	11	7	6	8	3	0	0	39	28	11
1997	4	13	6	8	12	6	0	0	49	31	18
1998	5	13	6	8	17	8	0	0	57	32	25
1999	6	12	9	13	14	10	3	2	69	40	29
2000	7	11	7	17	14	11	1	2	70	42	28
2001	7	11	8	18	12	11	1	1	69	44	25
2002	9	9	8	18	14	11	0	2	71	44	27
2003	13	5	14	13	13	10	1	1	70	45	25
Total	**146**	**173**	**97**	**112**	**132**	**85**	**6**	**8**	**759**	**528**	**231**
%	**19**	**23**	**13**	**15**	**17**	**11**	**1**	**1**	**100**	**70**	**30**

Ratings are observed on 5th March. Sovereigns with withdrawn ratings are eliminated from the year of withdrawal to the year of new rating. Default sovereigns are discarded from the year following default to the year of recovery. IG and NIG stand for Investment Grade and Non-Investment Grade issuers, respectively.

B. Percentage of rated sovereigns. Groups by rating and year

	Rating										
Year	Aaa	Aa	A	Baa	Ba	B	C	Default	Total	IG	NIG
1981	73	27	0	0	0	0	0	0	100	100	0
1982	73	27	0	0	0	0	0	0	100	100	0
1983	67	33	0	0	0	0	0	0	100	100	0
1984	67	33	0	0	0	0	0	0	100	100	0
1985	56	44	0	0	0	0	0	0	100	100	0
1986	75	25	0	0	0	0	0	0	100	100	0
1987	57	14	7	7	14	0	0	0	100	86	14
1988	39	33	6	6	6	11	0	0	100	83	17
1989	40	30	10	5	5	10	0	0	100	85	15
1990	32	32	18	5	5	9	0	0	100	86	14
1991	23	41	18	5	9	5	0	0	100	86	14
1992	17	43	17	4	9	9	0	0	100	83	17
1993	17	42	17	4	13	8	0	0	100	79	21
1994	17	30	20	3	23	7	0	0	100	70	30
1995	11	31	17	9	26	6	0	0	100	69	31
1996	10	28	18	15	21	8	0	0	100	92	28
1997	8	27	12	16	24	12	0	0	100	63	37
1998	9	23	11	14	30	14	0	0	100	56	44
1999	9	17	13	19	20	14	4	3	100	58	42
2000	10	16	10	24	20	16	1	3	100	60	40
2001	10	16	12	26	17	16	1	1	100	64	36
2002	13	13	11	25	20	15	0	3	100	62	38
2003	19	7	20	19	19	14	1	1	100	64	36

See note in Table A.

Appendix 5.4: Summary Statistics for Sovereign Rating Durations

Rating	Mean	StDev	Max	Min	StDev/Mean
Aaa	78.94	78.65	280.03	12.10	0.996
Aa	50.22	49.09	194.07	1.63	0.978
A	33.74	32.30	124.57	2.23	0.957
Baa	32.17	21.70	87.13	0.80	0.674
Ba	31.46	26.66	103.10	0.63	0.847
B	18.58	19.62	84.97	0.13	1.06
C	11.48	11.74	34.37	1.27	1.02
D	28.70	17.44	55.13	0.40	0.61

Duration is the number of months spent in each rating.

Data are from March 1981 to March 2004.