RATING HISTORY, TIME, AND THE DYNAMIC
ESTIMATION OF RATING MIGRATION HAZARD

By

Huong Dang, CFA, FRM

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Discipline of Finance, Faculty of Economics & Business at The University of Sydney

July, 2010
Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

The thesis contains no material previously written or published by any other person except where reference is made in the text of the thesis. I certify that all the assistance received in preparing this thesis and sources have been acknowledged.

(Signed)

Huong Dang
Dedication

This thesis is dedicated to my family without whose understanding and support my studies would have been impossible.
Acknowledgements

This thesis would not have been possible without my supervisor, Associate Professor Graham Partington, whose guidance, advice and support during the course of my doctoral studies, and the completion of this thesis, were invaluable. As well as dedicating his time and effort, his ideas and intellectual input were inspirational to me. He always stood by me and I could always rely on him for wise advice. He influenced not only my research interests but also my working methods. My gratitude is profound.

I would wish to thank Professor Paul Allison for his programming advice and helpful comments. I am indebted to Professor Tony Hsiu His Chen, Dr. Sam Li Sheng Chen and Dr. Amy Ming-Fang Yen for their programming assistance, helpful comments and funding support for my research trip to Taiwan in May, 2009.

Thanks are due to the participants of the following conferences, for their helpful comments: The Australian Banking and Finance Conference, 2007; The International C.R.E.D.I.T Conference on Credit Rating, 2007; The International Risk Management Conference, 2008 and 2010; The EFMA Conference, 2008; The European Risk Management Conferences, 2008 and 2009; The Paris International Meeting on Finance, 2008; The FMA European Conference, 2009; The INFINITI Conference on International Finance, 2009; The seminar at the Department of Economics and Finance, University of Canterbury, 2009; The seminar at the Department of Finance and Accounting, Macquarie University, 2009; The Finance and Corporate Governance Conference, 2010.

I am grateful to the Discipline of Finance, Faculty of Economics and Business at the University of Sydney, and the Capital Market Cooperative Research Centre (CMCRC) for their funding support during the course of my doctoral studies.

I am deeply grateful to my family for their constant support, encouragement and compassion. My family motivate me to strive to achieve and progress. Without all of their love and support, none of this would have been possible.

Lastly, my thanks are due to Stephen Edgar for his editing work, and to my friend, Robert Kelly, for his comments, assistance and support in times of need.
# Table of contents

Statement of Originality i
Dedication ii
Acknowledgements iii
Table of contents iv
List of Figures viii
List of Tables x
Abstract xi

## Chapter 1: Introduction
1.1. Introduction 1
1.2. Thesis objectives 4
1.3. Models and data employed 6
1.4. Summary of findings 9
1.5. Structure of the thesis 10

## Chapter 2: Literature review
Non-Markovian behaviours and time-heterogeneity in rating migrations
2.1. Markov property 14
2.2. Non-Markovian behaviours and time-heterogeneity 16
   2.2.1. Serial correlation 17
   2.2.2. Duration dependence 19
   2.2.3. Rating volatility 20
   2.2.4. Fallen angel events 21
   2.2.5. Rating cliff or substantial rating changes 24
   2.2.6. Rating withdrawals 26
   2.2.7. Original rating 27
   2.2.8. Rating age 28
   2.2.9. Industrial heterogeneity 30
   2.2.10. Time-heterogeneity 31
      2.2.10.1. Structural changes, changing rating policy and business cycles 31
      2.2.10.2. The evolution of time 34
   2.2.11. Upgrade and downgrade dynamics 36
2.3. Conclusion 38

## Chapter 3: Literature review
Methods to estimate rating migration hazards and to assess the predictive accuracy of probability forecasts
3.1. Estimation models 43
   3.1.1. Cohort method and transition matrices 44
   3.1.2. Qualitative response static models 46
   3.1.3. Survival analysis and Cox’s hazard model 48
References

Appendix A: The two sample t test and the Wilcoxon two sample test

Table A-1: Macro-economic time series
Table A-2: Proportion of upgrades/ downgrades
Table A-3: Time to events
Table A-4: Start rating
Table A-5: Rating history
Table A-6: Rating path
List of figures

Figure 4-1: Arrangement of rating states in the calendar time risk set. 65
Figure 4-2: Measurements of macro-economic covariates for estimation states in the calendar time risk set. 68
Figure 4-3: Arrangement of rating states in the event time risk set. 70
Figure 4-4: The construction of macro-economic covariates for holdout states. 81
Figure 4-5: The construction of the time interaction covariates for rating states in the estimation and holdout samples. 82
Figure 4-6: Covariance graph_Upgrade model, time=3. 89
Figure 4-7: Performance of the US's stock and money market 105
Figure 4-8: Performance of the US's economy 105
Figure 4-9: The distribution of rating changes and rating withdrawals by year of event, 1984-2000. 110
Figure 4-10: The distribution of rating changes and rating withdrawals by year of event, 2001-2005. 110
Figure 4-11: The distribution of rating changes and rating withdrawals by start rating 1984-2000. 111
Figure 4-12: The distribution of rating changes and rating withdrawals by start rating 2001-2005. 111
Figure 4-13: Histogram of downgrade changes, 1984-2000. 114
Figure 4-14: Histogram of upgrade changes, 1984-2000. 114
Figure 4-15: Histogram of time to events. 115
Figure 4-16: Histogram of start rating, 1984-2000 118
Figure 4-17: Histogram of start rating, 2001-2005

Figure 4-18: Distribution of rating states by the original rating.

Figure 4-19: Distribution of rating states in the estimation sample by the year first rated

Figure 4-20: Distribution of rating states in the holdout sample by the year first rated

Figure 5-1: Brier score covariance decompositions of survival forecasts generated by the stratified time-fixed downgrade model.

Figure 5-2: Brier score covariance decompositions of survival forecasts generated by the stratified time-fixed upgrade model.

Figure 7-1: Brier score covariance decompositions of survival forecasts generated by the stratified TVC downgrade model.

Figure 7-2: Brier score covariance decompositions of survival forecasts generated by the stratified TVC base upgrade model.

Figure 7-3: Brier score covariance decompositions of survival forecasts generated by the stratified TVC extended downgrade model.

Figure 7-4: The effect of dummy lag1 down on downgrade hazard in the TVC base and TVC extended models.

Figure 7-5: The effect of number NR on downgrade hazard in the TVC base and TVC extended models.

Figure 7-6: Brier score covariance decompositions of survival forecasts generated by the stratified TVC extended upgrade model.
List of tables

Table 4-1: Unconditional forecast horizons 80
Table 4-2: Conditional forecast horizons 84
Table 4-3A: Rating codes 93
Table 4-3B: Migrations either departed from or ended at top and bottom rating grades 95
Table 4-4: Descriptive statistics of macro-economic variables 106
Table 4-5: The construction of estimation and holdout samples 107
Table 4-6: Distribution of rating states by start rating and start year, 1984-2000 108
Table 4-7: Distribution of rating states by start rating and end rating, 1984-2000 109
Table 4-8: Descriptive statistics of time to event 116
Table 4-9: Descriptive statistics of start rating 117
Table 4-10: Descriptive statistics of rating history 119
Table 5-1: Stratified Cox’s proportional (time-fixed) hazard model 125-126
Table 5-2: Brier score of survival forecasts generated by the stratified time-fixed downgrade model 132
Table 5-3: Brier score of survival forecasts generated by the stratified time-fixed upgrade model 138
Table 6-1: Stratified time-varying covariate (TVC) base model and time-fixed model 146-147
Table 6-2: Stratified time-varying covariate (TVC) base and TVC extended model 151-152
Table 7-1: Brier score of survival forecasts generated by the stratified TVC base downgrade model 160
Table 7-2: Brier score of survival forecasts generated by the stratified TVC base upgrade model 165
Table 7-3: Brier score of survival forecasts generated by stratified downgrade models 170
Table 7-4: Brier score of survival forecasts generated by the stratified TVC extended downgrade model 171
Table 7-5: Brier score of survival forecasts generated by stratified upgrade models 179
Table 7-6: Brier score of survival forecasts generated by the stratified TVC extended upgrade model 180
Table 8-1: Substantial rating changes owing to multiple notch numeric rating scale 198
Abstract

This thesis employs survival analysis framework (Allison, 1984) and the Cox’s hazard model (Cox, 1972) to estimate the probability that a credit rating survives in its current grade at a certain forecast horizon. The Cox’s hazard model resolves some significant drawbacks of the conventional estimation approaches. It allows a rigorous testing of non-Markovian behaviours and time heterogeneity in rating dynamics. It accounts for the changes in risk factors over time, and features the time structure of probability survival estimates.

The thesis estimates three stratified Cox’s hazard models, including a proportional hazard model, and two dynamic hazard models which account for the changes in macro-economic conditions, and the passage of survival time over rating durations. The estimation of these stratified Cox’s hazard models for downgrades and upgrades offers improved understanding of the impact of rating history in a static and a dynamic estimation framework. The thesis overcomes the computational challenges involved in forming dynamic probability estimates when the standard proportionality assumption of Cox’s model does not hold and when the data sample includes multiple strata.

It is found that the probability of rating migrations is a function of rating history and that rating history is more important than the current rating in determining the probability of a rating change. Switching from a static estimation framework to a dynamic estimation framework does not alter the effect of rating history on the rating migration hazard. It is also found that rating history and the current rating interact with time. As the rating duration extends, the main effects of rating history and
current rating variables decay. Accounting for this decay has a substantial impact on
the risk of rating transitions. Downgrades are more affected by rating history and time
interactions than upgrades.

To evaluate the predictive performance of rating history, the Brier score (Brier, 1950)
and its covariance decomposition (Yates, 1982) were employed. Tests of forecast
accuracy suggest that rating history has some predictive power for future rating
changes. The findings suggest that an accurate forecast framework is more likely to be
constructed if non-Markovian behaviours and time heterogeneity are incorporated into
credit risk models.
Chapter 1

Introduction

1.1. Introduction

Credit ratings are widely used to assess the risk that an entity (i.e. a sovereign, a corporate) will default. Credit ratings can reflect the capability and willingness of an entity to meet a specific financial obligation (issue rating), or the overall creditworthiness of an obligor “to pay its financial obligations, incorporating an assessment of all future events to the extent they are known or can be anticipated” (issuer rating) (Vazza, Leung, Alsati, and Katz, 2005, p. 2). Credit ratings serve as important inputs for end-users in both the public and private sectors, including regulators, credit risk managers in financial institutions, investors, and companies themselves to make decisions on investment and risk management. In this thesis, rating migration or transition is a discrete process in which a corporate issuer moves from the current rating, namely the start rating, to a new rating grade, namely the end rating. A migration could, for example, be a one-notch downgrade from A to A- or a multiple-notch upgrade from B- to BB. A rating migration has a direct impact on the cost of capital of the issuer, on the portfolio investment performance of an investor, and on the regulatory risk capital requirements of a bank.

Accurately modelling credit risks based on rating migration probabilities has become an important issue since the new Basel Accord Framework (Basel II) came into effect. The Basel II framework allows banks to determine their capital adequacy requirement based on the credit quality of their counterparties. Since Basel II, the probability of rating migrations has been increasingly employed in pricing debt and in managing
credit risk. It is crucial to apply a suitable approach to estimating rating migrations and to model such migrations accurately, as a small change in the probability of a migration estimate may result in a large variation in regulatory risk capital requirements. For instance, Jafry and Schuermann (2004) demonstrated that employing different approaches such as discrete time cohort Markov, time-homogeneous / time-varying duration (continuous-time) to estimate rating migrations results in substantially different risk capital requirements. Switching across estimation approaches leads to more variation in the economic risk capital than switching between recession and expansion.

In recent years, credit rating agencies have been severely criticized for their slow responses to the deteriorating credit quality of corporate issuers and for the inaccurate assessment of the credit risk of structured products. An accurate estimation method requires an understanding of rating migration dynamics. Furthermore, an understanding of rating dynamics is useful to predict price movements in both bond and equity markets. Bond re-ratings can affect the price of equity securities as well as bonds, and price effects may spread to rivals (Dichev and Piotroski, 2001).

In practice, the discrete time cohort Markov framework is widely used by credit rating agencies to construct a rating migration matrix. This method derives the migration probability across different rating categories from the relative frequencies of rating changes observed in the past. The underlying assumption is that the migration probabilities follow a stable Markov property. Specifically, the probability that an issuer will migrate from its current rating to another rating category in this period is assumed to be independent of its rating behaviour in the past (Markov process), and remains constant over time (time-homogeneous).
There is current interest regarding whether the rating process truly is Markov and time-homogeneous. Taking a Markov approach to rating history, all of the relevant information about the impact of history is captured in the firm’s current rating (start rating) since it defines the beginning of its existing rating state. It is suggested that the Markov property adequately holds within a one or two-year horizon but does not persist at longer horizons (Kiefer and Larson, 2007; Frydman and Schuermann, 2008). There is also ample evidence that rating history variables are significant predictors of future rating distributions (see Altman and Kao, 1991; Altman and Kao, 1992a; Altman and Kao, 1992b; Carty and Fons, 1993; Altman, 1998; Kavvathas, 2001; Lando and Skodeberg, 2002; Hamilton and Cantor, 2004; Figlewski, Frydman, and Liang, 2008). The effect of rating history, however, diminishes or vanishes after a few years (Altman and Kao, 1991; Altman, 1992; Altman, 1998; Hamilton and Cantor, 2004; Fledelius, Lando, and Nielsen, 2004).

Evidence from previous studies supports the argument that the time-homogeneity property does not hold over the long term and there is a drift in migration mobility (Jafry and Schuermann, 2004; Kiefer and Larson, 2007). Issuers exhibit different time-varying migration patterns as time unfolds (Carty and Fons, 1993; Kavvathas, 2001; Koopman, Lucas, and Monteiro, 2006), and rating stability changes over time (Altman and Kao, 1991). The heterogeneity with respect to the rate of movement does not vanish after accounting for the state of the business cycle or the industry sector (Frydman and Schuermann, 2008).

The above evidence suggests that the transition probability vectors differ across issuers of the same rating grade, and vary across time. This emphasises the need to account for the rating history dependence (non-Markovian behaviours) and the time dependence in estimating rating migration probability. The following questions arise:
(i) whether and how rating history can explain time-varying rating migration probabilities; (ii) whether rating history interacts with time; (iii) if so, what is the nature of the time interactions; (iv) how the time interactions affect the main effects of rating history variables on subsequent rating changes.

A major part of this thesis aims to develop a model that captures time-heterogeneity in rating migration dynamics, and contributes further empirical evidence on non-Markovian behaviour by answering these questions.

It is important to emphasise that this thesis is about the effect of rating history and time on rating migrations. The thesis does not attempt to develop an optimal model that is predictive of rating change. That task is beyond the scope of this research.

However, assessing the forecast performance of rating history at various horizons may inform the development of such an optimal predictive model. Two natural questions then arise: (i) what is the predictive performance of rating history in forecasting time-varying rating migration probabilities?, and (ii) does the predictive accuracy change after accounting for the passage of time in the current rating grade? A significant part of this thesis addresses these questions.

1.2. Thesis objectives

Building on the conclusions drawn from chapter 2, which discusses the literature on credit rating migration dynamics, and chapter 3, which reviews the literature on approaches to derive rating migration probability and to assess the forecast performance of rating systems, this thesis has four main objectives.

First, motivated by the current debate on whether the rating process is stable Markov and the important role of credit ratings in current risk management practice, the thesis aims to examine a wide variety of non-Markovian behaviours in rating migration
dynamics. The question addressed is how rating history and the passage of time affect the probability of a rating downgrade and a rating upgrade. This question is examined in both static and dynamic estimation frameworks. The results support the argument that rating history significantly impacts on subsequent rating changes. Furthermore, its effect remains strong and intact after controlling for the dynamic evolution of macro-economic conditions.

Additionally, informed by the evidence that the effect of rating age and a rating drift vanishes as time extends (Altman and Kao, 1991; Altman, 1992; Altman, 1998; Hamilton and Cantor, 2004; Fledelius, Lando, and Nielsen, 2004), the thesis further investigates whether rating history variables interact with the duration of a rating state, and contributes substantial new evidence on the decay effect of rating history as a rating state continues in its current grade.

Second, motivated by the limitations of the methods commonly used to estimate migration probabilities and by the need to account for the changes in risk factors over time, this thesis develops stratified Cox’s hazard models for the hazard of a rating migration. The estimation approach accounts for the time sequence of recurrent migration events that each issuer experienced since the beginning of the observation period.

The thesis estimates three stratified Cox’s hazard models including a basic proportional hazards model with time-fixed covariates, and two dynamic models. Time-fixed covariates have a constant value, which is the value measured at the beginning of the rating state. The first dynamic model accounts for changing macro-economic conditions by using time-varying covariates (TVC base model). Time-varying covariates have values that are updated over the period an issuer remains in a rating grade. The second dynamic model allows for the interactions between the main
rating history effect variables and the time spent in a rating grade (TVC extended model). The estimation of these stratified Cox’s hazard models for downgrades and upgrades offers an improved understanding of the impact of rating history in a static and a dynamic estimation framework and after accounting for the effect of the passage of time.

Third, motivated by the need to incorporate the time dimension into probability forecasts, the thesis aims to generate forecasts of time-varying migration probabilities. The thesis overcomes the computational challenges involved in forming dynamic probability estimates when the standard proportionality assumption of Cox’s model does not hold (dynamic hazard model with time-varying covariates) and when the data sample includes multiple strata (stratified hazard model).

Fourth, motivated by the limitations of the methods commonly used to assess the forecasting power of credit rating systems, the thesis employs a proper probability scoring rule. The Brier Score (Brier, 1950) was chosen for this purpose since it can be decomposed into the components that contribute to forecast accuracy. The covariance decomposition (Yates, 1982) is used for this purpose, which has not been widely used in previous financial studies.

1.3. Models and data employed

Cox’s hazard model (Cox, 1972) is the method chosen to estimate the probability that a rating survives in its current grade at any point in time \( t \) over the time horizon of interest \( T \). The conventional Cox’s proportional hazard model (Cox, 1972) utilises time-independent covariates and a static estimation framework, which captures the values of the risk factors at a particular point in time. A positive feature of Cox’s
hazard model is that it is possible to embed the time dimension into the risk factors examined in the form of time-varying covariates. In the presence of such time-varying covariates, Cox’s hazard model employs a dynamic estimation framework that captures changes in the risk factors over time.

Three variations of the stratified Cox’s hazard model (Cox, 1972) were developed to model rating downgrades and upgrades. The stratified hazard model applies the conditional gap time approach, and accounts for the time sequence of repeated migrations each issuer experienced since the beginning of the estimation period (Prentice, Williams and Peterson, 1981). Observations of the same migration sequence are placed in the same stratum and have the same baseline hazard functions\(^1\).

Of the three variations developed, the simplest is the stratified Cox’s proportional model that employs time-independent (time-fixed) macro-economic covariates. The time-fixed model controls for the macro-economic conditions prevailing at the beginning of each rating state but does not account for the development of the macro-economic environment over time. It is likely that the credit risk profile of an issuer and its migration probability are more affected by the recent macro-economic conditions than those prevailing at the beginning of the current rating state. The time-fixed model was estimated to investigate the effect of rating history on migration hazard in a static framework. It was logical to subsequently examine non-Markovian behaviours in a dynamic estimation framework, and to develop a dynamic hazard model with time-varying covariates.

The stratified dynamic Cox’s hazard base and extended models incorporate the same set of rating history covariates and control variables as the stratified Cox’s

\(^1\) The baseline hazard is the hazard at time \(t\) if all of the variables (covariates) are set to a value of zero.
proportional hazard model. The difference lies in the construction of time-varying covariates. Both stratified dynamic models include time-fixed and time-varying macro-economic covariates. They both account for the macro-economic conditions at the commencement of each rating observation (time-fixed covariate) and macro-economic changes over time (time-varying covariates).

The dynamic extended model differs from the dynamic base model in that it incorporates additional time-varying covariates (TVC) that capture the interaction between rating history and time. The dynamic extended model thus accounts for the passage of time over rating durations. The migration hazard is then allowed to depend on the time spent in the current rating, and time is embedded in the interaction terms, which are updated whenever a migration event occurs in the estimation sample.

The thesis examines the probable duration of rating grades, that is how probable is it that by time $t$ (in the set $t = 1$ to $T$) there will be an upgrade or a downgrade? It is possible to derive this probability from a survivor function $S(t)$ which can be estimated using Cox’s proportional hazard model (Cox, 1972), where $S(t) = P(T>t)$ and $P(T>t)$ is the probability that the time of the rating transition $T$ will be after time $t$.

The probability estimate is dynamic as $t$ can take on a range of values. In the presence of time-varying covariates (TVC), generating time-varying probability forecasts from a dynamic Cox’s hazard model proved to be problematic. This is due to the computational challenges in estimating the baseline hazard function when the proportionality assumption does not hold. The issue becomes much more complicated for a data sample with multiple strata as each strata requires the estimation of a distinct baseline hazard function (stratified dynamic hazard model). To overcome these issues, the thesis applies an approach proposed by Andersen (1992) and extends
the work of Chen et al. (2005). Stratum-specific baseline hazard functions were estimated and time-varying probability survival forecasts were formed accordingly.

The forecast performance of rating history in a static estimation framework (time-fixed model) serves as the yardstick for comparative assessments when switching to a dynamic estimation framework (TVC base model) and accounting for the passage of survival time (TVC extended model).

The thesis employs Standard & Poor’s CreditPro2005 issuer rating data of US non-financial firms. In addition, US macro-economic data and industry sector categories were also utilised. The estimation sample covers the period 1984-2000, whereas the holdout sample (for out-of-sample forecast performance tests) covers the period 2001-2005.

1.4. Summary of findings

The key findings from the empirical studies carried out in this thesis are: (i) The evidence suggests that rating migrations are non-Markovian. A substantial number of rating history variables, in addition to the current rating, are significant in determining subsequent rating migration hazards. The rating history variables are more important than the current rating in explaining migrations. There is also some evidence that rating history tends to repeat itself. (ii) Downgrades and upgrades exhibit different migration dynamics, with downgrades being more affected by rating history and time. (iii) The impact of rating history and current rating variables decays as the rating duration expands and the time interaction dramatically changes the impact of some of the rating history variables. (iv) Overall, the forecast performance is disappointing. This may be partly explained by dramatic changes in conditions between the estimation and holdout periods. In the static estimation framework, rating history
exhibits a disappointing performance in predicting downgrades, and displays modest accuracy in forecasting short-term upgrade probabilities. (v) The dynamic estimation framework, which controls for changes in the macro-economic environment over time, improves the predictive performance of rating history in relation to short-term downgrades and longer-term upgrades. (vi) Taking into account the time spent in a rating grade improves the forecast accuracy of rating history; however, the improvement is strongly dependent on how informative the conditional survival duration\(^2\) of the holdout rating state is.

1.5. Structure of the thesis

The thesis is organised as follows.

Chapter 2 reviews relevant literature on the Markov property, non-Markovian behaviours and time-heterogeneity in rating dynamics. Key conclusions drawn from the literature review lay the foundation for the research questions that are subsequently examined in chapters 5 and 6.

Chapter 3 reviews the relevant literature on approaches to estimating rating migration probabilities and to assessing the predictive accuracy of probability forecasts. The first section of the chapter reviews the literature on the methods that are widely used to estimate the probability of rating changes, including the discrete time cohort Markov framework, and the qualitative response static models. It then reviews the literature on the survival analysis framework (Allison, 1984) and Cox’s hazard models (Cox, 1972). In addition, a brief review of the literature on methods to account

\(^2\) In order to form interaction terms with time in the holdout sample it has to be assumed that a rating survives until the date when the interaction term is computed. The conditional survival duration is the time \(T^*\) at which holdout observations were assumed to survive and their interaction terms were constructed.
for repeated migration events and sensitivity tests to detect informative censoring are also presented. The literature review on estimation models motivates the use of dynamic stratified Cox’s hazard models and the generation of time-varying probability forecasts in the presence of time-varying covariates.

The second part of chapter 3 reviews the literature on measures commonly used to evaluate the discrimination power of credit risk models including the ROC measure (Receiver Operating Characteristic) and the CAP measure (Cumulative Accuracy Profile). It then reviews literature on the Brier score (Brier, 1950) and approaches used to decompose the Brier score. This review motivates the use of the Brier score and the covariance decomposition (Yates, 1982) in evaluating the forecast performance of Cox’s hazard models developed in this thesis.

Chapter 4 presents the methods and data employed in this thesis. The first part of the chapter presents the arrangement of rating observations in event time risk sets, and the construction of time-fixed and time-varying covariates for estimation and holdout samples. It then presents the specifications of the stratified Cox’s hazard models (Cox, 1972) used in the thesis. It also presents the construction of benchmark (empirical frequencies) and naïve forecast models. The calculation of the Brier score (Brier, 1950) for time-varying probability forecasts, and the score covariance decomposition (Yates, 1982) for the estimated and benchmark models are also explained.

The second part of chapter 4 defines the rating history and macro-economic variables, industry control dummies, and interaction terms to be employed in the stratified Cox’s models. The third part of chapter 4 describes the construction of the estimation and holdout sample, and presents the descriptive statistics of rating observations in the estimation and holdout periods.
Chapter 5 presents the empirical results of the stratified Cox’s proportional hazard model (time-fixed model) and its forecast performance at various forecast horizons. The first part summarises the results of the estimated model with a focus on the impact of rating history on rating migrations. The second part of chapter 5 presents the Brier score results and the covariance decompositions of forecast survival probabilities at various horizons.

Chapter 6 presents the empirical results of the stratified Cox’s hazard models with time-varying covariates (TVC). The first section summarises the results of the stratified TVC hazard base model, which controls for the macro-economic conditions prevailing at the beginning of each rating state and the changes in macro-economic conditions over rating durations. The second section presents the results of the stratified TVC hazard extended model, which extends the TVC base model by employing additional time-varying covariates that capture the interactions between rating history variables and time.

Chapter 7 presents the Brier scores results and the covariance decompositions of time-varying probability estimates generated by the TVC hazard base and the TVC hazard extended models. The forecast performance of the time-fixed model developed in chapter 5 serves as the yardstick for comparative assessments across the models.

Chapter 8 summarises the key findings of the empirical research carried out in this thesis and discusses the practical implications of the findings. In addition, the chapter also discusses the limitations of the empirical studies and suggests directions for future research.
Chapter 2

Literature review

Non-Markovian behaviours and time-heterogeneity in rating migrations

This chapter reviews the relevant literature on rating migration dynamics. The standard approach to model rating migrations relies on two assumptions. First, the Markov property which assumes that prospective rating distributions depend on the current rating grade and that rating history is irrelevant in determining subsequent migration probabilities. Second, the time-homogeneous property which assumes that migration probabilities do not vary over time. The accuracy of these assumptions has been the subject of intense controversy.

The research literature has produced mounting evidence of non-Markovian behaviours such as rating momentum, duration dependence, rating age, time and industry heterogeneity in ratings migrations. In the light of the observations of path-dependent behaviours and time-heterogeneity in the rating process, it is natural to raise the question whether, and how, rating history and time can explain and predict time-varying rating migration hazards.

The objective of this chapter is twofold. The first objective is to review the literature on non-Markovian behaviours and time-heterogeneity in rating migrations. In order to provide a context for the literature on non-Markovian behaviors, the literature on the

---

4 Carty and Fons (1993), Lando and Skodeberg (2002)
5 Altman (1998), Figlewski et al. (2008)
Markov property in rating dynamics is reviewed first. The second objective is to identify research issues and develop the hypotheses that will be examined and tested in this thesis. The review of literature on rating migration dynamics leads to the key research objective, which is to determine the effect of rating history variables and their interaction with time on rating migration hazard.

The final part of the chapter sets out conclusions from the literature review, which lays the foundation for the research carried out in this thesis.

2.1. Markov property

The Markov property has been extensively used to model the rating migration process. Empirical studies found evidence that the current rating grade impacts on the stability of the rating process. For instance, issuers in high ratings are more stable than those in low ratings, and therefore less likely to experience defaults (Carty, 1997; Figlewski et al., 2008). In the same vein Hamilton and Cantor (2004, p. 7) indicated that controlling for past rating changes, “the default rates increase monotonically by whole letter rating categories”. Rating stability varies between investment rated and speculative rated issuers (Jorion, Shi, and Zhang, 2005). Hamilton and Cantor (2004) stressed that investment-grade issuers exhibit stronger rating stability following an upgrade than do speculative-grade obligors. Rating stability can be attributed to the fact that credit rating agencies focus on the long-term investment horizon (Altman and Rijken, 2004), and ratings are immune to cyclical fluctuations in credit quality (Altman and Rijken, 2004; Loffler, 2005).

The current rating also affects future rating distributions. For instance, issuers rated Aa or A exhibit a strong tendency for downgrades to exceed upgrades at any given time horizon (Carty and Fons, 1993, p. 12). Those in middle-rating categories have a
higher probability of drifting either way in the rating spectrum, and are more volatile. However, there is no consensus with regard to the migration propensity of low ratings. On the one hand, Lucas and Lonski (1992, p. 11) indicated that firms rated Ca and C are less volatile due to their proximity to the default state. By way of contrast, Carty and Fons (1993, p. 12) suggested that Caa rating is “too weak to make the uphill climb” and the rating shows a tendency to default.

The above evidence emphasises the need to control for the current rating grade in modelling rating migrations. Consistent with Carty (1997), this thesis hypothesises that the higher the current rating, the more likely that the issuer retains and stays in its current grade.

The literature also emphasises that obligors in the boundary of investment and speculative rating grades, compared with those in the high end and low end rating spectrum, show different migration propensities. For instance, issuers rated B have a greater likelihood of going up the rating scales than issuers in the boundary of speculative grade Ba (Carty and Fons, 1993; Carty, 1997). Bonds carrying high-end investment grades, compared with bonds rated BBB, are more likely to retain their current rating grades (Livingston, Naranjo, and Zhou, 2008).

However, there is mixed evidence on the rating process of issuers in the investment-grade boundary. On the one hand, Carty and Fons (1993), Carty (1997) showed that in the long term, obligors on the threshold of investment grade (Baa) are more likely to rise to higher rating grades than to fall to speculative grades. On the other hand, Johnson (2004) suggested that the lowest investment-grade rating, BBB-, is more likely to be downgraded than are its neighbouring ratings. He further stated that over a one-day period downgrades starting from a BBB- rating are likely to fall more grades than downgrades starting from the adjacent rating states. This pattern,
according to Johnson (2004), is understandable as there are few obligors for whom credit rating agencies will allow a significant default risk within their definition of an investment grade. Once these issuers lose their investment grade, they are subject to high default risk and thus likely to travel multiple notches downward. The ratings then change drastically and in quick succession rather than through gradual, mild adjustments. Standard & Poor’s (2001), however, suggested a different view. Substantial rating changes from BBB- may be possible simply because once issuers lose investment grade status their operations are impaired due to regulations or private contracts.

Given the above evidence, it is necessary to account for “the proximity” of the current rating to the investment and speculative grade boundary. This thesis conjectures that a rating currently in the investment or speculative boundary is likely to ascend the rating spectrum. Firms just above the boundary have a strong incentive to avoid a downgrade and firms just below the boundary have a strong incentive to work for an upgrade.

2.2. Non-Markovian behaviours and time-heterogeneity

The literature presents ample evidence of non-Markovian behaviours and time-heterogeneity in ratings migrations. For example, obligors of the same rating migrate at different rates, and the heterogeneity persists after controlling for the state of the business cycle or industry sector (Frydman and Schuermann, 2008). The sources of heterogeneity for issuers of the same rating grades include serial correlation/rating drift, duration dependence, rating volatility, fallen angel/rising star events, substantial rating jumps, original rating, rating age, industrial heterogeneity, and time variation due in part to business cycles. The following review presents the literature on non-Markovian behaviours and time-heterogeneity in rating dynamics.
2.2.1. Serial correlation

The direction of a prior rating change impacts on the departure probability out of the present rating grade (Hamilton and Cantor, 2004). Previous studies suggested that downgrades exhibit serial correlation (Carty and Fons, 1993; Altman and Kao, 1992a; Altman and Kao, 1992b; Kavvathas, 2001; Bangia et al., 2002; Lando and Skodeberg, 2002; Hamilton and Cantor, 2004; Mah and Verde, 2004; Figlewski et al., 2008). Specifically, issuers downgraded to a given rating, compared with those upgraded to that rating grade, are more likely to experience a subsequent downgrade. Positive autocorrelation is most evident for extreme rating grades (Altman and Kao, 1992b; Kavvathas, 2001). Rating dynamics of sovereign ratings also exhibit serial correlation, and a downgrade of multiple-notch ratings makes a further downgrade more likely than a single-notch downgrade (Al-Sakka and Gwilym, 2009). The evidence of serial autocorrelation when the initial rating change was an upgrade, however, is less obvious (Carty and Fons, 1993; Altman and Kao, 1992b).

The asymmetric effects of serial correlations were emphasised by Hamilton and Cantor (2004, pp. 6-7). Firstly, a recent downgrade has a stronger impact on the downgrade and default hazards than does a recent upgrade. Controlling for rating category, the risk of default “increases monotonically from lagged upgrade to lagged downgrade”. The probability of experiencing a subsequent migration in the same direction versus a migration in the opposite direction in the next year is nine times more likely for a downgrade and twice as likely for an upgrade occurring in the past year. Secondly, the effects of lagged downgrades on investment-grade-rated and speculative-grade-rated obligors are asymmetric at both one-year and three-year forecast horizons. Given a previous downgrade, investment-grade-rated issuers are 18 times more likely to be downgraded than upgraded within one year. On the other
hand, a downgrade makes speculative-grade-rated issuers six times more likely to descend than to ascend the rating spectrum over the same horizon. However, the strong effect of a previous rating change becomes weaker with the passage of time (Hamilton and Cantor, 2004, p. 10), or does not hold after two or three years (Fledelius, Lando, and Nielsen, 2004).

The evidence of rating drift, i.e. subsequent rating change in the same direction as the previous rating change, is consistent with credit rating agencies’ practice to reduce rating volatility and to be ex-post credible (Posch, 2006, pp. 1-2). The rating of an obligor will be revised if there is an enduring change in its fundamental credit risk, and the adjustment is unlikely to be reversed shortly afterwards. Posch (2006, p. 1) found that credit rating agencies will respond to changes in the fundamental credit risk of an issuer if default probability estimates change by around two notches. The timeliness of rating revision varies across rating territory and across time. Responses tend to be faster when default occurrences are high and when credit quality is low. In the same vein, Altman and Rijken (2004) stated that a rating change is triggered when the difference between the actual agency rating and the rating predicted by the agency rating model surpasses a certain threshold. Furthermore, if revised, ratings are only partly adjusted.

Rating drift can sometimes be attributed to credit rating agencies’ attempts to “dole out the bad news in small doses rather than savaging the bond issuer all in one go” (Economist, 1997 in Loffler, 2005, p. 374). Stylised facts, such as rating stability, serial correlation, and ratings lag changes in issuers’ default risk, are also consistent with the policy of rating bounce avoidance. Loffler (2005, p. 366) proposed that there exist tolerance regions around rating boundaries, and rating revision will not occur if credit quality crosses a boundary but still lies within the tolerance area. In other
words, rating are revised “when the credit quality is relatively close to the boundary triggering a further rating change”, and “rating changes are suppressed rather than handed out piecemeal” (Loffler, 2005, p. 374). This is not surprising as credit rating agencies act in the interests of their clients, i.e. bond issuers, to avoid costly frequent reversals, especially for downgrades.

The finding of serial correlation in rating dynamics raises the question of how rating history can explain and predict subsequent rating changes. Consistent with the evidence discussed above, this thesis hypothesises that there is a tendency for rating history to repeat itself and that recent history has the strongest effect. If this is so, the direction of the immediately prior rating regrades will positively affect the probability of a further regrade in the same direction and negatively affect the probability of a regrade in the reverse direction.

2.2.2. Duration dependence

There is mixed evidence on the effects of lagged rating duration on subsequent rating changes. On the one hand, Carty and Fons (1993, p. 20) indicated that the longer a rating is maintained, the more likely it is that a rating change will occur. Furthermore, the length of time an issuer stays in a particular rating grade (survival duration) positively correlates with its long-term credit quality. For instance, an Aaa rating is expected to “survive” for 10 years whereas a B-rated issuer has an expected lifetime of four years.

On the other hand, Lando and Skodeberg (2002, pp.437-440) demonstrated that the migration probability negatively correlates with the length of time an issuer retains its current rating. They suggested that the negative duration effect and downward momentum reflect the practice of credit rating agencies to revise rating grade by one
notch at a time and through a “series of mild downgrades”. This is consistent with the conjecture that ratings tend to change gradually in response to the growth or decline of credit risks (Carty, 1997). In the same vein, Hamilton and Cantor (2004, p. 3) affirmed that Moody’s tend to “limit rating reversal and dampen rating volatility” by revising ratings “in a gradual, even predictable fashion”. Furthermore, rating revisions are often done “incrementally in the same direction rather than with a single, multiple-notch rating change”, according to Hamilton and Cantor (2004, p. 5).

The above results support the need to control for duration dependence in modelling rating migrations. Based on the evidence provided by Lando and Skodeberg (2002), this thesis hypothesises that the longer the duration of a directly prior rating state, the longer the likely duration of the current rating state. Any duration effect may, however, be conditional on the regrades being in the same direction. The disruption in continuity created by regrades in opposite directions could disrupt persistence in duration. The effect of recent history on duration is expected to diminish with the passage of time and if so the penultimate (lag-one) duration will have a stronger impact on the migration hazard than the antepenultimate (lag-two) duration.

The evidence of a duration effect also motivates the use of Cox’s hazard model, since an attractive feature of this model is that it controls for the underlying duration dependence, without requiring specification of the functional form of that dependence.

2.2.3. Rating volatility

An increase in rating volatility over time, with downgrades outnumbering upgrades, was documented by Altman and Kao (1991), Lucas and Lonski (1992, p. 7), Lando and Skodeberg (2002). Koopman et al. (2006) observed high re-ratings and default
hazards for firms with very short survival durations in the low-investment ratings, and attributed this property to rating momentum. Lando and Skodeberg (2002, p. 437) suggested that issuers who, on their downward journey, stay a short time in the middle rating grades are likely to descend to the lower rungs of the rating territory.

The above evidence highlights the need to account for downgrade volatility and rating change volatility in estimating rating migration hazards. This thesis conjectures that a history of frequent rating downgrades is likely to be repeated. That is, the higher the rate of prior downgrades the higher the hazard of subsequent downgrades. It is not certain whether a history of frequent rating changes makes a further migration more likely. However, given the asymmetric effects of downgrade and upgrade (Hamilton and Cantor, 2004), it is hypothesised that downgrade volatility dominates rating change volatility.

2.2.4. Fallen angel events

Given the dependence of hazard rates on rating history, it is relevant to investigate whether different rating paths lead to different migration hazards.

The rating process of fallen angels, which were initially investment-grade rated but experienced deterioration in credit quality and fell to speculative-grade ratings, has been of particular interest in the literature\(^7\). Empirical studies affirm some perceptions about fallen angels, and support the notion that fallen angels exhibit different rating dynamics compared with their peers. For instance, Figlewski et al. (2008) observe

---

\(^7\) The migration dynamics of fallen angels bear significant market impacts on the cost of borrowing, financial strategy flexibility, and portfolio compositions. Investment grade bond funds, subject to restricted or targeted risk level, may not be able to hold non-investment grade bonds or they may be limited to holding a small percentage of speculative grade bonds such as fallen angels (Altman and Kao, 1992b, p. 73).
different hazards of rating changes for fallen angels, rising stars, and those that maintain broad investment/speculative rating categories.

The rating a fallen angels receives, at the time they are downgraded from investment to speculative grade, has a strong impact on their default probability as well as their probability of becoming a rising star⁸ (Mann, Hamilton, Varma, and Cantor, 2003, p. 5). Fallen angels which received a Ba rating, compared with other fallen angels, exhibit a higher propensity to rise to the investment grade territory and are less vulnerable to default. The lower the rating a fallen angel descends to, the more likely it will default. The higher the rating a fallen angels receives prior to and after its fall date, the more likely it will become a “prodigal son” and return to the investment grade spectrum.

Within the first few years since falling to speculative grades, fallen angels exhibit strong downward momentum (Mann et al., 2003; Vazza, Aurora, Schneck, 2005). For instance, 49 percent of defaulted fallen angels, compared with 36 percent of their defaulted peers⁹, experience credit deterioration in the initial three year period of financial difficulties (Vazza et al., 2005, p. 12). Fallen angels, compared with their peers, experience a faster rate of migration from the date they lost investment grade status till the date they fall to their lowest rating grades (Vazza et al., 2005). A faster downward migration makes a fallen angel more vulnerable to default risk. A variety of causal factors that can be attributed to the greater rating velocity of fallen angels include increasing competitive pressures and changing industry dynamics, lack of

---

⁸ Rising stars are cases which were upgraded from speculative rating grades to investment rating grades.

⁹ Vazza, Aurora, Schneck (2005) defined fallen angel peers as those which were originally rated speculative grades and have identical rating distribution characteristics.
access to short-term capital, higher leverage strategy, failed leverage buy-outs, unwise acquisition, and financial distress.

During the initial years since their fall, fallen angels are a riskier proposition than their peers\(^\text{10}\), as evidenced by their higher likelihood to default and shorter median time to default (Mann et al., 2003; Vazza et al., 2005). However, fallen angels, on average, reach slightly lower rating at five years as compared to the rating they received on the date they fell (Mann et al., 2003). This suggests that fallen angels rebound strongly after surviving the initial period of financial difficulties.

In the long run fallen angels are less risky and, relative to their peers, more likely to rise back to investment grade territory if they survive the initial years of high risk and overcome immediate obstacles (Mann et al., 2003; Vazza et al., 2005). Fallen angels display a greater tendency to survive, and non-defaulted fallen angels seem to cling to life for many more years than their counterparts. This outcome can be attributed to the characteristics of fallen angels and their counterparts. Compared with their peers, fallen angels have the “advantaged debt structures” of larger issues. They can issue debt with favourable terms such as lower coupon, longer maturities and fewer restrictive covenants (Mann et al., 2003, p. 2). As time passes, surviving fallen angels possess robust franchise value, enhanced business strength, and improved profitability (Vazza et al., 2005). On the other hand, their peers generally come from new industries with high business risks, or follow financial policies that target lower ratings (Vazza et al., 2005, p. 11).

\(^{10}\) Mann, Hamilton, Varma, and Cantor (2003) defined fallen angel peers as speculative-grade rated issuers that were of the same ratings as the fallen angels at the time they lost investment grade status, and never rated in the investment grade spectrum.
Among fallen angels, utilities are more likely to be upgraded for every horizon whereas non-utilities do not have exceptional performance or a favourable experience towards upgrades (Altman and Kao, 1992a, p. 19). Furthermore, the magnitude of subsequent rating changes for non-utilities fallen angels was “dramatic”. Public utility fallen angels exhibit strong negative serial correlation, that is, a downgrade to speculative ratings is more likely to be followed by an upgrade. On the other hand, non-utility fallen angels exhibit positive serial correlation, and are likely to descend to the lower ends of the rating scale after losing their investment-grade ratings.

Fallen angels in high-velocity sectors such as telecommunications exhibit rapid deterioration in credit profile but they show strong recovery within five years of their fall date (Vazza et al., 2005, p. 17). This result is intuitive as within high-velocity sectors defaulted fallen angels tend to be weeded out during shakeouts while resilient fallen angels cling to life and regain a higher average rating five years after their fall date. By way of contrast, fallen angles in low-velocity sectors such as leisure/media experience a slow decline in creditworthiness after losing investment grade status, but they do not make a strong rebound in the subsequent years.

Consistent with prior research (Mann et al., 2003; Vazza et al., 2005), this thesis hypothesises that having been a fallen angel makes a further upgrade more probable. This thesis also examines whether rating migration hazard is affected by being a rising star, or by a history of substantial jumps across rating grades.

### 2.2.5. Rating cliff or substantial rating changes

A substantial rating change of multiple notches, known as a rating cliff, reflects a substantial decline or improvement in the credit risk of an issuer. Substantial rating changes of more than one letter grade were more frequently observed in the ratings B
through C during the period 1970-1990 (Lucas and Lonski, 1992, p. 12), but were less frequent than rating revisions of small magnitude (Carty and Fons, 1993; Carty, 1997). This result is consistent with Moody’s policy to “keep large magnitude rating changes” to a minimum (Carty and Fons, 1993, p. 10).

A rating cliff may happen for a number of reasons and may, or may not, raise a question regarding the accuracy of the original rating (Standard & Poor’s, 2001). In some situations, as certain scenarios develop, issuers are likely to experience default or a substantial deterioration in credit quality. The issuers are then prone to a rapid downward spiral, rather than a series of gradual downgrades. For example, issuers rated BBB-, the lowest investment grade, are sensitive to rating triggers. Once downgraded to speculative ratings, their costs of capital soar. Rating triggers combined with other factors may then accelerate the process by which their cost of capital substantially increases, due to their deteriorating credit quality. Significant liquidity issues then beset these fallen angels, particularly when multiple triggers occur simultaneously (Standard & Poor’s, 2001).

By way of contrast, a credit cliff may merely reflect an “unusual sensitivity to credit quality of a particular occurrence” (Standard & Poor’s, 2001). Examples include government subsidised dependencies (i.e. ailing Californian utilities which ultimately did not receive the needed support from the State that they initially expected), confidence-sensitive and capital-intensive entities (i.e. troubled financial institutions), rating-triggered situations that require more collateral or accelerated loan repayments if the rating of the issuer falls below a certain threshold, structured finance, insured ratings, and catastrophe bonds.

Given the variety of situations that are associated with celebrated downgrades, it is not clear whether a history of substantial rating changes repeats itself or whether
rating dynamics exhibit mean reversion. The mean reversion propensity was documented for middle rating grades (Altman and Kao, 1992b), and for low-end investment grades (Kavvathas, 2001). Specifically, a prior upgrade for those rated BBB or A makes a subsequent upgrade less likely. This implies that most obligors attain an “average rating” under normal circumstances (Kavvathas, 2001, pp. 32-33). Issuers initially rated in the middle of the rating scale do not exhibit a tendency to substantially drift in either direction, and ratings also tend to migrate toward the middle of the rating spectrum (Altman and Kao, 1992b, p. 70).

Based on the evidence presented above, this thesis conjectures that a history of substantial downgrades (upgrades) makes a subsequent upgrade (downgrades) more likely.

2.2.6. Rating withdrawals

A period of being unrated (NR) creates a break in rating history. Observed results suggest that rating withdrawals are mostly associated with the issuer’s exit from the public bond market. In other words, most obligors withdraw from being rated because they no longer carry significant debt. Of all observed rating withdrawals, 95 percent were not associated with a decline in creditworthiness, four percent occurred for unspecified reasons and only one percent were associated with an increase in credit risk (Carty, 1997, p. 10).

The reasons behind a rating withdrawal are varied and include a lack of information to accurately assess the debt issue, the retirement of the underlying issue, early redemption of the issue due to conversions or mergers and bankruptcies. In the case of withdrawal because of a lack of information to accurately assess the debt issue, rating withdrawal bears negative credit implications (Carty, 1997, p. 10). Issuers are likely
to withdraw from being rated when they expect downgrades. Where a decline in the credit profile is only apparent to the issuers they may decide to bypass credit rating agencies. These firms may then choose to be re-rated when they are likely to receive a better credit rating. Alternatively, firms that have become NR may be obliged to seek re-rating when they need to make a debt issue, even if their rating has not improved. It is not clear what a period of being unrated signifies about current credit quality, or the probability of rating regrades.

Rating withdrawals account for 2-3 percent of the number of issues after one year and 25-40 percent after five years (Altman, 1998, p. 1235). This observation can be partly attributed to the fact that small obligors with lower ratings tend to replace rated public bonds with more attractive unrated private debts. As the creditworthiness of an issuer deteriorates the probability of rating withdrawal increases (Carty, 1997, p. 10). This thesis hypothesises that if firms repeatedly use becoming unrated as a strategy to avoid downgrades, then their probability of being downgraded will be reduced.

2.2.7. Original rating

An obvious question to ask is for how long is the impact of rating history felt? The answer is that it extends right back to the firm’s first rating (original rating). The literature provides evidence that issues of different original rating grades exhibit different rating migration dynamics (Altman and Kao, 1991; Altman and Kao, 1992b; Jorion et al., 2005; Figlewski et al., 2008).

New issues of different ratings retain their original ratings in a different fashion and their rating stability varies over time (Altman and Kao, 1991; Altman and Kao, 1992b). Issues originally rated AAA/ AA/ A that are downgraded tend to suffer a subsequent downgrade, and a majority experienced a negative first rating change for
all horizons (Altman and Kao, 1992a, p. 18). The difference between downgrades and upgrades increases as the time horizon expands (Altman and Kao, 1992b). Within the first five years post-issuance, AAA-rated bonds are most stable in retaining their initial rating. However, as the time horizon expands to ten years, A and B-rated bonds exhibit higher stability than those rated AAA. BB-rated bonds, which are in the boundary of the speculative rating spectrum, show the least stability in retaining the initial rating over time and do not display a clear propensity to migrate in either direction. On the other hand, BBB-rated bonds, which are in the boundary of investment ratings, show a predominance of upgrades over downgrades (Altman and Kao, 1991 pp. 19-20; Altman and Kao, 1992a; Altman and Kao, 1992b, pp. 65-67).

Originally rated investment-grade bonds generally are more likely to be downgraded than upgraded, whereas bonds that were originally assigned speculative-grade ratings exhibit neither a propensity to descend nor ascend the rating spectrum (Altman and Kao, 1992a, p. 16). If an originally issued high-yield bond is downgraded, a further downgrade is more likely. The reverse also applies. If an originally issued high-yield bond is upgraded, the next rating change is more likely to be an upgrade than a downgrade.

Given the above evidence, the original rating is incorporated in the models used in this thesis. The question is whether its effect remains in the presence of the additional rating history variables that the model also includes.

**2.2.8. Rating age**

The time since first rating may also be expected to impact on rating migrations. Altman and Kao (1991), Altman (1992), Altman (1998), Figlewski *et al.* (2008) demonstrated that a newly rated issue, compared with a seasoned issue of the same
rating grade, has a smaller probability of default. The aging effect is more relevant for seasoned bonds and those of low credit ratings. For instance, seasoned junk bonds are more likely to be downgraded in one year than newly issued bonds of high credit quality (Altman and Kao, 1991). However, the aging effect vanishes after three or four years (Altman and Kao, 1991, p. 25; Altman, 1992, p. 89; Altman, 1998, p. 1234).

The aging effect for corporate bonds can be attributed to several factors. First, the longer the time since the bond was issued, the more likely the issue is called. The original population of bonds is then “reduced by call, sinking funds, and maturations, as well as by defaults and exchanges” (Altman, 1992, pp. 85-86). Also, issuers of high credit quality are much more likely to call their bonds or to repurchase their public rated debts in the open market at attractive prices than those of poor credit quality. The pool of remaining bonds therefore becomes more vulnerable to default than the original universe (Altman, 1992, p. 88). Second, firms with newly issued bonds generally receive the face value of the issuance and have sufficient working capital to service their debts. These firms tend to exhibit lower credit risk compared to firms with seasoned bonds (Altman, 1998, pp. 1239-1240). Third, firms are generally under “intense credit rating scrutiny” at the time they issue public rated debts (Altman and Kao, 1991, p. 25). Following the issuance, credit reviews are conducted less frequently and not for “at least one year” in most cases. It is unlikely that a default would occur during the initial year, though a decline in creditworthiness could develop long before a credit review takes place. Altman (1998, pp. 1239-1240) indicated that a rating revision could occur to a newly issued bond only if a substantial deterioration in credit quality is imminent. On the other hand, seasoned bonds could have experienced gradual decline in credit quality and their ratings would
be revised after some time. As a result, a new issue may retain its original rating for a longer period of time than a seasoned issue. Altman and Kao (1991, p. 26) further stressed that “the intensity and timeliness of rating scrutiny” on new issues and seasoned issues can affect “the reliability of rating drift results encompassing bonds of different ages and time periods”.

As this thesis focuses on the effects of rating history, observations that pass screening tests are those which experienced at least two migrations (i.e. they have non-censored lag-one and lag-two duration data). The estimation sample therefore consists of seasoned issuers. In light of the literature on the aging effect of seasoned issuers, this thesis hypothesises that the longer the period of time since an issuer was first rated the more likely it is to be downgraded.

2.2.9. Industrial heterogeneity

The literature offers evidence of industrial heterogeneity in rating migration dynamics. For instance, for the ten-year horizon, industrial issuers of investment grades are more stable in retaining their original rating (Altman and Kao, 1991). For non-investment-grade obligors, those in industrial sectors are slightly more stable, especially during the five-year holding period (Altman and Kao, 1991). Industrial issuers exhibit lower rating volatility than banks but are vulnerable to multiple-notch rating changes as much as, or more than, banks (Nickell et al., 2000).

There are different industry sector sensitivities to state factors and hence different migration dynamics across industries (Kavvathas, 2001). In the same vein, Kadam and Lenk (2008) suggested that rating migration behaviours vary across industry

---

11 The mean rating ages of issuers in the estimation and holdout samples are 8.578 years and 10.257 years respectively. Altman and Kao (1992a) reported that during the period 1970-1985, it took up to four years for the majority of issuers to experience their first rating migrations.
sectors; for instance the ratings of utility sector issuers are more stable than those of industrial sector obligors. This finding can be attributed to the fact that the volatility of future revenues varies across industry sectors.

Industry risk constitutes an important determinant of the assigned rating as each industry faces an upper-limit rating (Galil, 2003). Berd (2005, p. 23) stressed that “distinguishing between broad industry sectors” improves the explanatory power of rating migrations more than taking into account a few macro factors. He obtained the best-fitting model by combining macro-economic and industry factors. Consequently this thesis incorporates industry controls in the models to capture the industry risk of each issuer.

2.2.10. Time-heterogeneity

2.2.10.1. Structural changes, changing rating policy, and business cycles

The time-homogeneity property assumes that rating distributions are constant over time. The literature, however, documents that rating stability changes over time. Corporate creditworthiness, as evidenced by rating changes, was less stable in the 1980s than in the 1970s (Altman and Kao, 1991, pp. 19-20). The occurrence of downgrades in the 1980s was far greater than in the 1970s. This pattern is evident in most of the rating grades. For instance, the 1970s saw 20 percent of AAA-rated issues fall to lower ratings, whereas the 1980s observed more than half of them fail to retain their original rating. BB-rated bonds displayed a tendency for upgrades to exceed downgrades in the 1970s but the proportion of upgrades is about the same as that of downgrades in the 1980s. The migration pattern for CCC-rated issues reversed across the two decades, with downgrades surpassing upgrades in the 1980s. Only BBB-rated bonds show a consistent propensity for upgrades to predominate over downgrades.
across the two decades. However, the difference between upgrades and downgrades substantially decreased in the 1980s (Altman and Kao, 1991).

Investment-grade ratings deteriorate over time, according to Blume, Lim and MacKinlay (1998). The cause that leads to the deteriorating quality of investment-rated issuers is a matter of controversy. On the one hand, the decrease in average rating of investment-rated issuers was attributed to a shift away from less risky sectors, higher volatility of stock prices, lower quality of accounting information, and increased earning management (Jorion et al., 2005). On the other hand, Blume et al., (1998, p. 1412) suggested that credit rating agencies applied more stringent rating standards over time “in terms of the explanatory variables used in the analysis”. This view implied that the observed decrease in average rating in the early 1990s is the result of tightening rating methodology rather than a decline in the actual credit quality of US corporate debt over time. However, the fact that Blume et al. (1998) neglected speculative firms could result in bias in their estimates as low-rated firms tend to be sensitive to business cycles and might be “subject to intensive monitoring” in recessions (Amato and Furfine, 2003).

Previous studies also found that ratings move pro-cyclically. The rating universe evolves differently in periods of contraction as opposed to expansion, (Bangia et al., 2002). Periods of economic contraction see a greater occurrence of downgrades and defaults, whereas periods of high growth observe more incidences of upgrades (Nickell et al., 2000). In addition, rating volatility decreases during business cycle peaks and increases during business cycle troughs (Nickell et al., 2000). Furthermore, periods of weak real GDP growth see an increased generation of fallen angels, whereas fewer incidences are observed in times of strong GDP growth (Vazza et al., 2005). It is also clear that the time taken to migrate to default decreases during
economic downturns (Helwegen and Kleiman, 1997; Belkin, Suchower, Forest, 1998; Alessandrini, 1999; Kavvathas, 2001; and Duffie, Saita, and Wang, 2007). These findings are not surprising, as the evolution of economic cycles drives the correlations in credit risks of issuers across industries, as Wilson (1997) observes.

It is obvious that business cycles impact on rating migration dynamics (Bangia et al., 2002; McNeil and Wendin, 2006; Kadam and Lenk, 2008; and Figlewski et al., 2008). For instance, a lower equity return and a higher equity return volatility makes a further downgrade more likely, according to Kavvathas (2001). Blume, Keim and Patel (1991) stressed that rating migrations are primarily impacted by macro-economic conditions rather than individual characteristics of public-rated debts. The significant systematic risk factors vary across transitions, and downgrades involve different systematic risk factors compared to upgrades (Koopman, Kraussl, Lucas, and Monteiro, 2009). In addition, firms with different ratings tend to react to the same economic conditions in different ways (Altman and Kao, 1992b).

The sensitivity of credit ratings to the stages of the business cycle can be attributed to the over-optimism/ pessimism of credit rating agencies in revising ratings during economic expansions/ contractions (Amato and Furfine, 2003). The state of the economy should be incorporated in credit risk models as it is a major driver of systematic credit risk, and low rating is more sensitive to macro-economic conditions than high rating. Failure to account for the state of the economy could result in an underestimation of “downward potential of high yield portfolio” in contraction periods or “suboptimal capital allocation in lending business” (Bangia et al., 2002, p. 469). The above findings emphasise the need to control for economic conditions in the models.
2.2.10.2. The evolution of time

The common understanding is that the time-homogeneity property may be applied to estimate rating migrations over a one-year period (Jafry and Schuermann, 2004). In the same vein, Kiefer and Larson (2007) indicated that Standard & Poor’s corporate ratings are time-homogeneous Markov over three transitions but this property breaks down at four migrations. This implies that the time-homogeneous Markov property does not hold over the long run, and care should be taken when estimating rating migrations for more than a year or two (Kiefer and Larson, 2007, p. 833).

Previous studies have found ample evidence of time-heterogeneity in the rating process. For instance, issuers of different rating grades exhibited different time-varying migration patterns as time unfolds (Carty and Fons, 1993, p. 18). Those of high credit quality, rated Aaa or Aa, have a migration probability that increases over time. Issuers with the highest rating Aaa can, of course, only maintain their rating or descend the rating scales. Their credit quality therefore gradually deteriorates over a long horizon and they travel downward at an increasing pace over time. On the other hand, obligors of low credit ratings, such as B and Caa, have a transition probability that decreases with time. These obligors are vulnerable to a rating downgrade in the short-term period. If they survive but fail to substantially improve their credit quality, they tend to maintain their existing rating categories over a long horizon. Thus, their migration probability decreases as the time horizon expands. Those in the middle of the rating scales, such as A, Baa and Ba, can either go up or go down the rating spectrum. Their probability of a rating change therefore tends to vary at a constant rate as time progresses.

Similarly, different patterns in the hazard rates for a downgrade and an upgrade were observed as the time horizon expands (Kavvathas, 2001, pp. 26-27). Upgrades,
particularly for issuers in low investment grades, exhibit increasing hazard rates over time, whereas this tendency is less prominent for downgrads, especially those with lower credit ratings. Of particular interest, CCC-rated obligors have hazard rates that decrease as time passes. This propensity is not surprising as low-rated issuers, if they survive the initial distress time, tend to cling to life and become less risky (Carty and Fons, 1993, p. 18).

Different time-varying rating processes for issuers in investment and sub-investment grades were also documented by Koopman *et al.* (2006). For instance, short-lived obligors in the sub-investment rating grades are vulnerable to defaults. Those clinging to life for a longer period of time tend to exhibit high default intensities at the time they have to roll over initial debts. The peak in defaults was observed after three to five years. By way of contrast, issuers in the investment-grade spectrum are most at risk of default during the first two years and less likely to default thereafter.

The findings described above rule out the assumption of time-homogeneity in rating process. Therefore, it is necessary to control for the passage of time in rating dynamics. Koopman *et al.* (2006) addressed the time-heterogeneity in rating migration intensity for firms rated at investment and speculative grades by constructing the baseline hazard as a function of the time spent in the current grade.

This thesis takes a different approach to account for the time an obligor has retained the current rating grade. While the rating continues in its current state the distance in time from the historic observations is extending. The impact of historic variables on the migration probability are thus likely to become increasingly “stale”. The point here is not just that more distant variables are likely to be less relevant, but that the impact of the variables interacts with the duration of the rating state. If so, interaction variables for rating history with time should reduce the impact of the main effect
variables for history. Based on the evidence that the effect of rating age and a rating drift vanish over time (Altman and Kao, 1991; Altman, 1992; Altman, 1998; Hamilton and Cantor, 2004; Fledelius et al., 2004), this thesis conjectures that the effects on transition probabilities, of rating history and the current rating, decay as the current rating continues through time.

2.2.11. Upgrade and downgrade dynamics

Numerous studies highlight that rating changes affect stock and bond returns in an asymmetric fashion. A possible explanation for the asymmetric effects of rating changes is that downgrades may be more difficult to forecast than upgrades and may follow different stochastic processes.

The previous rating grade and previous rating event correlate with the magnitude of underperformance in both stock and CDS markets (Norden and Weber, 2004). There are differential stock price responses to upgrades and downgrades (Holthausen and Leftwich, 1986; Hand, Holthausen, and Leftwich, 1992; Norden and Weber, 2004). While upgrade announcements have no impact, downgrade announcements are associated with abnormally low returns. Dichev and Piotroski (2001) found a negative abnormal stock return varying from 10 percent to 14 percent within the first year following a downgrade, and attributed this result to the overreaction of the stock market to rating downgrades. Only downgrades associated with a decline in credit risks result in an underperformance on the stock market (Goh and Ederington, 1993). This pattern, however, does not hold for downgrades associated with leverage or restructure causal reasons. The impact of downgrades on stock reactions vary across speculative and investment rating grades (Goh and Ederington, 1999). Specifically, downgrades to and within speculative rating grades result in more pronounced negative stock reactions.
The asymmetry of price adjustment has also been observed in the bond market. Wansley, Glascock, and Clauretie (1992) observed the strong negative impact of downgrades on bond returns before and after the downgrade announcement. They further indicate that underperformance in the bond market is correlated with the magnitude of rating downgrades. Downgrades to and within speculative-grade rating lead to more pronounced bond reactions (Hite and Warga, 1997; Dynkin, Hyman, Konstantinovski, 2002). However, upgrade announcements do not result in any effect on bond returns, according to Wansley et al. (1992), and Hite et al. (1997).

Prior studies also support the view that downgrades and upgrades are driven by different risk factors. The impact of the common (systematic) risk factors is higher for downgrades than for upgrades, and varies across different migration routes (Koopman et al., 2006). Upgrades depend more on firm-specific shocks than do downgrades (Kavvatha 2001; Koopman et al., 2006).

Consistent with the evidence that downgrades and upgrades are driven by different risk factors (Kavvatha, 2001; Koopman et al., 2006), and downgrades exhibit rating momentum (Carty and Fons, 1993; Altman and Kao, 1992b; Kavvathas, 2001; Lando and Skodeberg, 2002; Hamilton and Cantor, 2004; Mah and Verde, 2004; Figlewski et al., 2008), it is hypothesised that the rating process of downgrades and upgrades differs, with downgrades depending more on rating history and time.

In the light of the foregoing evidence on dissimilar behaviours, this thesis develops separate models for upgrades and downgrades. This approach has been widely applied in previous studies such as Lando and Skodeberg (2002), Livingston et al. (2008), Al-Sakka and Gwilym (2009).
2.3. Conclusion

From the review of literature on the Markov property, non-Markovian behaviours, and time-heterogeneity in rating dynamics, the following conclusions may be drawn:

(i) Rating dynamics are assumed to follow the time-homogeneous Markov property. Empirical studies found that within one or two years the Markov property adequately holds (Jafry and Schuermann, 2004; Kiefer and Larson, 2007).

(ii) There is mounting evidence that rating dynamics exhibit non-Markovian behaviours. A variety of rating history aspects affect subsequent rating distributions. Furthermore, the time-homogeneous property does not hold over the long run (Kiefer and Larson, 2007), and rating migration hazards vary over time.

(iii) The current rating affects subsequent rating distributions (Lucas and Lonski, 1992; Carty and Fons, 1993), and high ratings are more stable than low ratings (Carty, 1997; Figlewski et al., 2008). The evidence of rating stability is consistent with through-the-cycle rating policy (Altman and Rijken, 2004; Loffler, 2005).

(iv) Issuers rated in the boundary of investment and speculative rating grades exhibit a different migration tendency relative to issuers in the high end and low end rating territory (Carty and Fons, 1993; Carty, 1997; Johnson, 2004; Livingston et al., 2008). However, there is no consensus with regard to the specific migration dynamics of issuers in the boundary rating thresholds.

(v) Rating downgrades exhibit serial correlation (Carty and Fons, 1993; Altman and Kao, 1992a; Altman and Kao, 1992b; Kavvathas, 2001; Bangia et al., 2002;
Lando and Skodeberg, 2002; Hamilton and Cantor, 2004; Mah and Verde, 2004; Figlewski et al., 2008), and this is most obvious for extreme rating grades (Altman and Kao, 1992b; Kavvathas, 2001). A lagged downgrade has a stronger impact on the downgrade and default hazards than a lagged upgrade (Hamilton and Cantor, 2004). Furthermore, a lagged downgrade affects investment-grade rated and speculative-grade rated issues in an asymmetric fashion (Hamilton and Cantor, 2004). However, the strong effect of lagged rating change decays as time unfolds (Hamilton and Cantor, 2004; Fledelius et al., 2004). The evidence of serial correlation is consistent with credit rating agencies’ policy of rating bounce avoidance (Loffler, 2005) and their practice of diminishing rating volatility (Altman and Rijken, 2004; Posch, 2006).

(vi) The duration of lagged rating change affects future rating migration hazard (Carty and Fons, 1993; Lando and Skodeberg, 2002). The negative duration effect on subsequent rating change is consistent with credit rating agencies’ policy to “dampen rating volatility”, “limit rating reversal” (Hamilton and Cantor, 2004), and to revise rating grades “by one notch at a time” (Lando and Skodeberg, 2002).

(vii) Rating volatility increases over time with downgrades more frequently observed than upgrades (Lucas and Lonski, 1992; Lando and Skodeberg, 2002). Issuers who stay a short time in the low investment ratings/middle rating grades are more likely to travel down than up the rating spectrum (Koopman et al., 2006; Lando and Skodeberg, 2002).

(viii) Different rating paths lead to different migration hazards (Figlewski et al., 2008). The occurrence of a fallen angel event affects rating migration probability (Mann et al., 2003; Vazza et al., 2005). Fallen angels exhibit a
stronger propensity to default and are less likely to regain investment-grade status than their counterparts within the first few years of falling to speculative grades. However, if they survive the initial years of financial distress, fallen angels are resilient, compared with their peers, and more likely to rise back to investment-grade territory.

(ix) Substantial rating changes were less frequently observed than one-notch rating changes, and were more frequently observed in the low-rating categories. Substantial rating revisions occur in a number of scenarios, which may or may not lead to a substantial deterioration in credit quality and default (Standard & Poor’s, 2001). It is uncertain whether a history of rating cliff repeats or rating mean reversion dominates.

(x) Rating withdrawals are mostly associated with the issuer’s exit from the public bond market, and occasionally bear negative credit implications. It is not clear what impact an unrated period may have on the probability of rating migrations, once the firm becomes rated again.

(xi) The length of time since an issuer was first rated also affects rating migrations (Altman and Kao, 1991; Altman, 1992; Altman, 1998; Figlewski et al., 2008). Newly rated issues, relative to seasoned issues of the same rating, are less likely to be downgraded and default within the first few years. The aging effect is most obvious for low rating grades (Altman and Kao, 1991), and vanishes after three or four years (Altman and Kao, 1991; Altman, 1992; Altman, 1998).

(xii) Issuers of different original ratings exhibit different migration dynamics (Altman and Kao, 1991; Altman and Kao, 1992b; Jorion et al., 2005). Investment-grade issues are more likely to be downgraded than upgraded,
whereas issues that were original given speculative-grade ratings exhibit neither a tendency to travel up or down the rating scales (Altman and Kao, 1992a).

(xiii) There are different industry sector sensitivities to state factors (Kavvathas, 2001), and the volatility of future revenues varies across industry sectors (Kadam and Lenk, 2008). Each industry faces an upper limit rating (Galil, 2003), and the rating process exhibits industry heterogeneity (Altman and Kao, 1991; Nickell et al., 2000).

(xiv) Ratings move pro-cyclically (Nickell et al., 2000). Downgrades and defaults were more frequently observed in periods of economic contraction, whereas upgrades occur more often in periods of expansion. In addition, rating volatility decreases during business cycle peaks and increases during business cycle troughs (Nickell et al., 2000).

(xv) The time-heterogeneity in rating dynamics does not vanish after controlling for the business cycle or industry sector (Frydman and Schuermann, 2008). Rating stability changes over time (Altman and Kao, 1991). Furthermore, issuers of different rating grades exhibit different time-varying migration patterns (Carty and Fons, 1993).

(xvi) Rating changes affect stock and bond performance in an asymmetric fashion, with downgrades having a stronger impact than upgrades (Holthausen and Leftwich, 1986; Hand et al., 1992; Norden and Weber, 2004; Dichev and Piotroski, 2001; Wansley et al., 1992). Furthermore, downgrades to and within speculative grades lead to more pronounced stock and bond negative performance (Goh and Ederington, 1993; Hite and Warga, 1997; Dynkin et al.,
This implies that downgrades follow different stochastic processes and are more difficult to model than upgrades.

Downgrades and upgrades are driven by different risk factors. Downgrades are more sensitive to systematic factors whereas upgrades depend more on firm-specific factors (Kavvatha, 2001; Koopman et al., 2006). Generally, rating migrations and defaults are primarily impacted by macro-economic conditions (Blume et al., 1991).

The above conclusions set out the foundation for the research carried out in this thesis. Building on that foundation, subsequent chapters examine the following questions:

(i) Whether the current rating alone is significant in explaining rating migrations or whether additional rating history variables are significant?

(ii) Which variables are significant when a comprehensive set of rating history variables is considered, what is their effect, and how important are they relative to the start rating?

(iii) Whether rating history variables interact with time, and if so, which interaction terms are significant, what are the interaction effects, and how important are they relative to the main effects of rating history variables on migration hazard?

(iv) How different are the main effects of rating history variables and their interaction with time on downgrade and upgrade hazards?

(v) Whether rating history variables can predict subsequent rating changes, and if so how accurate are the time-varying probability forecasts?
Chapter 3

Literature review

Methods to estimate rating migration hazards and to assess the predictive accuracy of probability forecasts

This chapter examines the conventional approaches used to estimate rating migration hazards and to assess the predictive accuracy of probability forecasts. The objective of the chapter is twofold. The first part reviews the relevant literature on the estimation models commonly used in practice and in academic studies, followed by the introduction of a method that resolves some of the problems of the conventional models. The review then continues to one of the motivations for the research carried out in this thesis, which is the development of dynamic stratified Cox’s hazard models and the generation of time-varying probability forecasts in the presence of time-varying covariates. The second part of the chapter reviews the literature on methods commonly used to assess the forecasting performance of rating systems. An alternative method, the Brier score (Brier, 1950), and its covariance decomposition (Yates, 1982) is proposed. This can be used to evaluate various attributes of probability forecasts, such as calibration, discrimination and scatter.

3.1. Estimation models

Two approaches have been widely applied to derive the probability of rating migrations. The indirect approach, namely the discrete time cohort Markov method, has been used extensively in practice (Lucas and Lonski, 1992; Carty and Fons, 1993; Carty, 1997; Hamilton and Cantor, 2004). Under this approach, migration matrices
are constructed from historical migration rates, and the probability of subsequent rating migrations is derived from the migration matrices.

The direct approach uses static qualitative response models, such as logistic and probit regressions. Such models provide an estimate of the probability of a rating change for each issuer. Static qualitative response models have been used widely in the academic literature (Ohlson, 1980; Nickell et al., 2000; Blume et al., 1998).

### 3.1.1. Cohort method and transition matrices

The discrete time-homogeneous Markov process and the cohort framework have been applied extensively by credit rating agencies (Lucas and Lonski, 1992; Carty and Fons, 1993; Carty, 1997; Hamilton and Cantor, 2004) to estimate future rating distributions. The discrete time cohort method, however, has been under criticism in the academic literature for a number of reasons.

First, the approach ignores the exact timing and survival duration of each issuer in the cohort. The parameters obtained in the discrete time framework are also sensitive to the time units employed in the estimation procedure (Flinn and Heckman, 1982). By focusing on ratings at two points in time, the beginning of a period and the end of a period, transition matrices fail to take into account the intervening credit events (Carty and Fons, 1993, p. 10). In practice, it is annual transition matrices that are typically published by credit rating agencies. The discrete time yearly framework does not capture the duration and transition information arising during the year (Kavvathas, 2001).

Second, Bangia et al. (2002, p. 456) suggest that “as the ongoing coverage follows at least a quarterly review pattern transition matrices estimated over short time periods best reflect the rating process”. Shorter intervals capture more rating changes.
Multiple-notch migrations tend to take place via intermediary states, rather than as one big jump. Thus shorter intervals more accurately capture the migration dynamics. However, transition matrices constructed from the cohort method are not ideally suited to model rating distributions over the short-term horizon, as the diagonal terms, which show the proportions of rating unchanged, are close to unity while other terms, which represent rating changes, are close to zero (Johnson, 2004).

Third, the sparsity of data for rare events such as defaults from AAA results in unrealistic migration estimates (Lando and Skodeberg, 2002). It is difficult to calibrate the model for unobserved events in the conventional cohort framework. Hamilton, Ou, Kim, and Cantor (2007) noted that there was no direct default occurrence for investment-grade rated obligors during the period 1980-2006. In fact, none of these issuers is default risk free and they should therefore receive non-zero default probability estimates.

Fourth, the cohort method assumes that obligors of the same rating have the same credit quality, and ignores heterogeneity across issuers. For instance, by pooling issues outstanding at any given point in time in the basket, the average annual (discrete time) method ignores the aging phenomenon – that is that the probability of default increases as the bonds get older (Altman, 1992, p. 85). Indeed, the discrete time cohort approach assumes that the default likelihood of a newly issued bond is the same as that of a seasoned bond. Pooling together issues of various ages conceals the results for bonds at any given age (Altman and Kao, 1991, p. 25). Rating migrations also exhibit downward momentum. Obligors within the same (yearly) cohort can enter

---

12 Enron Corporation was rated investment grade till four days prior to its bankruptcy filing in 2001. WorldCom was rated investment grade until 42 days before its bankruptcy filing in 2002 (Johnson, 2004).
the cohort by either ascending from a lower rating or descending from a higher rating grade. Unconditional rating transition rates, which ignore the rating history of individual issuers, can be “misleading measures of individual issuer transition probabilities” (Hamilton and Cantor, 2004, p. 5).

Last but not least, the cohort method does not properly account for censoring. Censored cases are those issuers observed for a certain length of time but who have not experienced the event of interest under the study. What happens to issuers in the cohort after the sample period ends is unknown.

For the above reasons, the discrete time cohort method, though widely used due to its simplicity, is not an ideal approach.

**3.1.2. Qualitative response static models**

In the academic literature, the approaches commonly used to model credit events such as bankruptcy and rating downgrades include multiple discriminant analysis (Altman, 1968), and qualitative response static models such as logistic (Ohlson, 1980) and probit regression analysis (Nickell et al., 2000; Blume et al., 1998). These static models allow an estimate of the posterior probability that a firm with a given set of characteristics for a specific year will survive or become bankrupt at some point in time, but fail to give any insight on when the bankruptcy event will take place (Whalen, 1991).

Like the cohort method, the qualitative response models ignore the event time and the period at risk of each firm in the estimation process (LeClere, 1999). Over a long sampling period, lumping together firms that could have been at risk for different durations of time prior to going bankrupt clouds the results for firms of any given duration. These static models use cross-sectional data, take into account the value of a
variable at a point in time, and consider the firm’s profile at an instant of time (LeClere, 1999, p. 10). In the same vein, Shumway (2001, p. 102) indicated that the static models only consider one set of data for each observation, i.e. in the year prior to the event of bankruptcy, and do not consider “data on healthy firms that eventually go bankrupt”. For instance, Altman (1968) developed separate static models one year before bankruptcy, two years before bankruptcy, three years before bankruptcy and so on. Each model utilises data observed at a specific time prior to the event. Since the time to observe data was chosen arbitrarily, the estimates are subject to selection bias (Shumway, 2001, p. 102).

In addition, the single-period classification models (static models) fail to accommodate the fact that characteristics of firms and macro-economic conditions change over time (Shumway, 2001). These static models therefore explain little about the dynamics of credit risks, and do not shed light on the underlying process that generates the duration of time until an event of interest occurs.

Furthermore, static logistic and probit regression analysis can not extract information from censored observations (LeClere, 1999, p. 5). The qualitative response static models treat observations that leave the healthy group for reasons other than the event (bankruptcy) as healthy states rather than censored states (Shumway, 2001). Previous bankruptcy forecasting models, for instance by Altman (1968) and Olson (1980), are misspecified, according to Shumway (2001, p. 104). He claims that these static models are inappropriate for forecasting bankruptcy, as they generate predictive probabilities that are biased and inconsistent estimates of the actual probabilities. As a result, test statistics of static models provide incorrect inferences and are invalid (Shumway, 2001, p. 111).
Finally, the static models focus just on the observations of event/ non-event state. They are not equipped to give a warning of failure with sufficient lead time (LeClere, 1999, p. 38). Static logistic and probit regression analyses do not provide a forecast as a function of time and therefore fail to feature the time dynamic of survival/ failure probabilities. The need to take the time dimension into survival forecasts and to use covariate information that varies over time has been increasingly emphasised in finance studies during the past ten years.

3.1.3. Survival analysis and Cox’s hazard model

3.1.3.1. Survival analysis

The survival analysis framework (Allison, 1984), particularly Cox’s hazard model (Cox, 1972), has been increasingly used to model the hazard of credit risk events such as default and rating migrations. In so doing, it resolves the problems of the qualitative response static models and the discrete time cohort Markov method. The advantages of using the survival analysis framework to model credit events have been illustrated by LeClere (1999), Kavvathas (2001), Shumway (2001), Lando and Skodeberg (2002), Guttler (2009). The application of survival analysis in modelling credit events is appealing, as “financial distress does not occur instantaneously” (LeClere, 1999, p. 67). Indeed, the credit profile of firms gradually erodes by a series of mild downgrades, and generally deteriorates over a period of time before the bankruptcy event actually takes place.

Survival analysis utilises information from both event observations and censored observations in the estimation process (Guttler, 2009). According to Allison (1995), this feature results in consistent parameter estimates. Lando and Skodeberg (2002) supported the use of complete information and the full story of rating transitions
observed in the sample. Issuers that leave the sample period due to reasons unrelated to the event should be included in the data set and treated as censored observations.

Moreover, survival analysis sheds light on the time dynamic of survival forecasts (Banasik, Crook, and Thomas, 1999). This framework allows the estimation of the time to migration in addition to the migration/survival probabilities at different forecast horizons. The ability of survival analysis to generate time-varying forecasts is particularly important for risk management in financial institutions. For instance, regulators need to be informed about distressed financial institutions with sufficient lead time to make a decision on preventive or remedial actions (Halling and Hayden, 2006). Regulators also need to regularly examine the survival profile of each financial institution and evaluate its survival probability at various time horizons. In addition, time-varying probability survival forecasts serve as useful inputs for financial institutions to set efficient credit risk pricings and to make timely adjustments to regulatory risk capital requirements.

Furthermore, the survival analysis framework provides a convenient way to study non-Markovian behaviours, the effects of business cycles and changes in rating policy on migration dynamics (Lando and Skodeberg, 2002). For instance, hazard models can account for both duration dependence and firm age (Shumway, 2001).

Lastly, survival analysis can be applied to either discrete time or continuous time data. The framework also allows the inclusion of both time-fixed and time-varying covariates. LeClere (1999, pp. 3, 10) stated that survival analysis utilises “longitudinal data from the time period preceding bankruptcy”, and “incorporates changes in the covariates over time in the estimation process”. The framework can examine the impact of time-varying covariates on “the duration of time preceding the event” and “the probability that the event will occur” (LeClere, 1999, pp. 3). It places emphasis on
the substantive processes that govern the occurrence and timing of events (Allison, 1984).

3.1.3.1.1. Non-independent censoring

Survival analysis uses data from both the event and the censored observations in the estimation process. Censoring may occur for many reasons. For instance, firms may withdraw from being rated if they no longer carry any public debt. In this case, censoring is random and non-informative. According to Allison (1995, p. 12-13), random censoring occurs when observations leave the study for a reason that is not under the control of the investigators. Random censoring does not introduce any bias into the parameter estimates.

The critical issue is whether the reason that the censored observations leave the study is independent of the event of interest (LeClere, 1999, pp. 7-8). It is plausible to suppose that a firm of poor credit quality is likely to withdraw from being rated to bypass credit rating agencies and avoid an imminent downgrade in the near future. For instance, if a firm is rated speculative and its credit quality is deteriorating, the firm may choose to replace public rated debt with private debt and become not rated (censored). In such circumstances, becoming unrated might substitute for being downgraded, and this type of censoring could lead to informative censoring.

Informative censoring occurs when censored observations constitute a biased sub-sample of any non-censored observations with the same covariate values (LeClere, 1999, p. 65). Informative censoring introduces bias into parameter estimates. Unfortunately, there is no statistical test to check for informative censoring and no standard methods for handling informative censoring (Allison, 1995, p. 14).
Two sensitivity tests have been suggested by Allison (1995, pp. 249-252) to examine the effects that informative censoring has on the results. These two tests are based on two assumptions representing the worst case scenarios. First, censored firms are assumed to be of poor credit quality and would have experienced a downgrade immediately after leaving the study. Second, censored firms would have stayed in the study as long as any other firm in the study. A comparison of the results of the two sensitivity tests with the result of the base scenario would then reveal whether censoring is informative. If the results are not substantively different from the base scenario, the censoring can then be assumed random (Allison, 1995).

The sensitivity tests for informative censoring under two extreme assumptions (Allison, 1995) were applied to the Cox’s proportional hazard model developed in chapter 5. By varying the assumed duration of 646 not rated (NR) cases in the estimation sample, this thesis found that treating NR firms as either non-informative censored or non-censored makes no significant change to the results. It therefore appears that censoring by being unrated is not informative in this thesis. Furthermore, excluding NR cases from the estimation sample does not qualitatively change the results. The subsequent chapters present the results of the hazard models that include NR cases.

**3.1.3.1.2. Methods to account for repeated migration events**

Four popular procedures have been developed to account for recurrent events in modelling the hazard functions of repeated events (Hosmer, Lemeshow, and May, 2008, pp. 288-296). All approaches assume that censoring is independent of the incidence of events. The following discussion briefly reviews the alternative approaches and identifies the method used in this thesis.
The simplest approach, proposed by Andersen, Borgan, Gill and Keiding (1993), applies the counting process framework to handle multiple event data. In this approach, recurrent migration events of the same issuers are assumed to be independent. Dependence across rating states of the same issuer is treated by “adjusting the estimates of the standard errors” (Hosmer et al., 2008, p. 288). The model does not differentiate migration events of different sequences. Time is defined from the commencement of the study. In this approach a rating state that is observed at a particular time will contribute to the risk set\textsuperscript{13} formed at that time (Hosmer et al., 2008, p. 293). This means that a risk set could include a rating firm that has not experienced its first migration and also a rating firm that has undergone a number of migrations. Intuitively, the migration dynamics of a firm experiencing only one rating migration and a firm that has experienced a number of rating migrations are likely to be different. The counting process framework therefore seems inappropriate for this study.

The marginal event-specific model was developed by Wei, Lin, and Weissfeld (1989). In this setting, the clock is reset each time the subject experiences an event. Thus duration is measured from the time the subject enters each recurrent event. “Each event is analyzed as a separate process”, and “the total time to each of possible recurrent events is modelled” (Hosmer et al., 2008, pp. 290, 294). Rating states across migration sequences contribute to a risk set as long as they are under observation at the time the risk set is formed, regardless of the number of events actually experienced. This approach is also inappropriate to this study since it assumes each firm is simultaneously at risk of repeated events. For instance, a firm

\textsuperscript{13} A risk set constructed at time $t$ is composed of all the firm ratings that are at risk of a rating change at time $t$. 

can be at risk of a third migration even though it has not yet experienced the first and the second migration events.

Two conditional models proposed by Prentice, Williams and Peterson (1981) take into account the time sequence of recurrent migration events. These models assume a rating state will not be at risk of a migration event until it has undergone all prior events. The data are stratified by the number of migration events the firm has experienced. The stratum variable is used in the estimation process to keep track of the time sequence in repeated migration events since the commencement of the study. A firm moves to stratum $s$ immediately following its $(s-1)^{th}$ migration and remains there until the $s^{th}$ migration occurs or it becomes censored.

The difference between the two conditional models proposed by Prentice et al. (1981) lies in the starting time. The conditional counting process method defines time from the commencement of the study. In this model, rating states are arranged in calendar time risk sets. The baseline hazard function is a function of the time from the beginning of the study. On the other hand, the conditional gap time method models the gap times between repeated events. Time begins at zero as a rating state commences and ends when the state either migrates to another rating grade or becomes censored. In this model, rating states are arranged in the event time risk set formed at each event time. The baseline hazard function is a function of the gap time between successive migration events.

Given that the objective of this study is to investigate the impact of rating history on the survival duration (gap time) of rating states, the conditional gap time method (Prentice et al., 1981) is the logical choice.
3.1.3.2. Cox’s hazard models

Hazard models are the key to survival analysis. The attraction of hazard models is well articulated by Shumway, 2001 (pp. 102-103). Hazard model “resolves the problem of static models”, “gives consistent estimates”, produces “more precise parameter estimates” and “more efficient out of sample forecasts” by utilising richer data than static models. Hazard models avoid the selection bias, and “often produce dramatically different statistical inferences” than qualitative response static models. Furthermore, it is a “refined approach” to “address the real meaning of rating and overcome the curse of rare events” (Galil, 2003, p. 5).

One of the major techniques used in hazard model is Cox’s proportional hazard model (1972). Cox’s model is considered a pre-eminent method to estimate survival models, and has been widely used in biometrics, engineering, sociology, criminology, and has had some use in studies of bankruptcy14.

Cox’s proportional hazard model offers several advantages in modelling credit events such as rating migrations. It does not make any assumptions about the distribution of survival times (Allison, 1995, p. 183). The model does not require the precise measurement of the event times, only the time ranks are needed. In addition, the model can be adapted to model competing risks and accommodate a variety of migration routes, for instance, upgrade, downgrade, default, migration from speculative to investment grade or vice versa. Despite the name, the technique can be used to model non-proportional hazards. There is some loss of statistical power in the

Cox’s model, as no assumption is made about the parametric form of the hazard function. The compensation is that the model is more robust. It is found that Cox’s hazard model “identified failed and healthy banks with a high degree of accuracy” and “flagged a large proportion of banks that subsequently failed as potential failures in periods prior to their actual demise” (Whalen, 1991, p. 21).

The finance literature using Cox’s hazard model has used two approaches, using time-fixed variables and time-varying variables. The studies using time-fixed variables allow the generation of time-varying survival forecasts and out-of-sample forecast performance assessments. However, these proportional hazard models typically employ data one year or two years prior to the time of event, and model the survival function without accounting for the changes in the variables over time. In other words, the estimation of the Cox’s proportional hazard model does not account for the evolution of firm-specific and macro-economic factors over time. For instance, Lane et al. (1986), and Chen and Lee (1993) employed time-independent variables to analyse the bankruptcy process. These proportional hazard models could be “potentially mis-specified since the values of the covariates [variables] are likely to change as a firm approaches financial distress” (LeClere, 1999, p. 60).

Research that uses time-varying variables in Cox’s hazard models addresses the above issue but, as discussed in chapter 4, the hazards are no longer proportional in such models. Nevertheless, Cox’s model is particularly convenient for estimating hazard models with time-varying variables (Allison, 1995, p. 183).

15 See for example, Lane et al. (1986), Whalen (1991), Chen and Lee (1993), Lunde et al. (1999), Partington and Stevenson (2001), Yao, Partington and Stevenson (2005), Chncharat et al. (2007).

16 See Figlewski et al. (2008), Guttler (2009)
When modelling credit risk events, for instance bankruptcy or rating migrations, it is necessary to account for changes in financial data, which provide an indication of the firm’s deteriorating credit profile prior to the event (Shumway, 2001). The dynamic hazard models utilise each firm’s time series data “by including annual observations as time-varying covariates”; therefore they “explicitly account for time” and capture the time-varying risk of the event (Shumway, 2001, p. 102). On the downside, the presence of time-varying covariates makes the task of estimating the baseline hazard function, a key element in forming time-varying probability forecasts, challenging. Therefore, making forecasts with a dynamic Cox’s hazard model has been problematic. Figlewski et al. (2008), for example, used Cox’s hazard models with time-varying covariates to investigate the effects of dynamic macro-economic risk factors on default and migration hazards, but did not make any forecasts using their models.

The finance literature has been rather silent on developing a dynamic Cox’s hazard model and forming probability forecasts in the presence of time-varying covariates. In a recent medical study, Chen, Yen, Wu, Liao, Kuo, and Chen (2005) took an approach proposed by Andersen (1992) to estimate the integrated baseline hazard, and were able to predict the dynamic hazard of death by liver cancer in the presence of time-varying risk factors. The development of the dynamic hazard model with out-of-sample forecast performance in the medical field motivates this thesis to fill the gap in the finance literature.

To answer the questions raised in chapter 2, this thesis develops stratified Cox’s hazard models for upgrades and downgrades, with and without time-varying covariates. The model with time-fixed variables utilises rating history variables and macro-economic data measured at the beginning of each rating state. As argued
above, it is desirable to incorporate time-varying variables into the model. In order to accomplish this, the thesis modifies the SAS codes published by Chen et al. (2005) and extends their work to develop stratified Cox’s hazard models with time-varying covariates. In the dynamic Cox’s hazard model, the values of macro-economic covariates and the interactions between rating history and time were updated at each event time as obligors were followed during the study. The incorporation of time-varying covariates offers an opportunity to explore the dynamic associations between risk drivers and the hazard of the event (Banasik et al., 1999).

Similar dynamic Cox’s hazard models were developed by Figlewski et al. (2008). However, there are major differences between this study and the study by Figlewski et al. (2008). First, this thesis focuses on the effects of rating history and time on rating migration hazard, whereas Figlewski et al. (2008) placed their emphasis on the impacts of macro-economic conditions on the hazards of default and migrations. Second, this thesis considers every rating revision, which includes both minus and plus changes in rating grades, and examines generic migrations such as downgrades and upgrades. On the other hand, Figlewski et al. (2008) merged ratings into broad classes such as the investment class, and investigated specific migration events and defaults. Third, Figlewski et al. (2008) did not form probability forecasts and therefore were not able to evaluate the model’s out-of-sample forecast performance.

To my knowledge, this thesis is the first study that develops dynamic stratified Cox’s hazard models with time-varying firm-specific (interaction terms) and macro-economic covariates to generate dynamic probability forecasts of repeated rating migrations. A very positive feature is that this framework is well suited to developing dynamic models for the rating process. The dynamic estimation framework accommodates changes in the rating history of firms, particularly the time a firm spent
in its current rating state (survival duration), and the evolution of macro-economic conditions over time. The estimation of the stratified Cox’s hazard models specifically accounts for the sequence of repeated migration events that issuers experienced since the beginning of the observation period. The dynamic predictions allow the survival probabilities of issuers to vary over the forecast horizons.

3.2. Methods to assess probability forecast accuracy

“The important yardstick of success for a failure prediction or early warning model is its out-of-sample forecasting accuracy”, Whalen (1991, p. 28). The objective of this thesis is not to develop an early warning model, though the techniques applied and the results of this thesis could inform such a development. However, it is of interest to determine whether rating history does have predictive power for future rating migrations. This section, therefore, briefly discusses the approaches commonly used to assess the accuracy of credit risk models and introduces the Brier score (Brier, 1950), which is a pre-eminent method used to assess probability forecasts in meteorology.

3.2.1. Methods to assess the discrimination ability of forecasts

Prior studies\textsuperscript{17} devoted to assessing the predictive accuracy of credit risk models tend to emphasise discriminatory power rather than calibration ability. The two measures that are most commonly used to evaluate the discriminatory power of credit risk models and rating systems are the ROC curve (Receiver Operating Characteristic) and CAP curve (Cumulative Accuracy Profile) (Engelmann, Hayden and Tasche, 2003; Basel Committee on Banking Supervision (BCBS), 2005; Tasche, 2008). The area under a ROC curve and the accuracy ratio obtained from a CAP curve are equivalent measures and are connected by means of linear transformation (Engelmann \textit{et al.}, 2003). The two measures reflect the ability of estimated models to discriminate \textit{ex}

\textsuperscript{17} See, for example, Hamilton and Cantor (2004), Vazza, Leung, Alsati, and Katz (2005), Wong, Partington, Stevenson, and Torbey (2007), Metz and Donmez (2008).
Discrimination and calibration are quite different. For example, consider a forecast sample where 80 percent of the ratings turned out to be unchanged. If a model gave the probability of a rating migration as 0.51 for all ratings that migrated and 0.5 for all ratings that remained unchanged, it would have perfect discrimination, but poor calibration. Calibration or reliability reflects the ability of estimated models to “assign the right numerical label” to probability forecasts (Yates, 1982, p. 150-151). Bessler and Ruffley (2004, p. 399) indicate that “calibration is a test of whether an issued probability agrees with its relative frequency, *ex post*”. Correct calibration of a rating system is equivalent to accurate default estimates, and calibration quality is “an important prudential criteria to assess rating system” (BCBS, 2005, p. 29). In order to have the capacity to examine both discrimination and calibration, this thesis assesses forecast accuracy using the Brier score (Brier, 1950). This is because the Brier score can be decomposed to reveal various component parts of forecast performance.

### 3.2.2. Brier score

The Brier score (Brier, 1950) is used in this thesis to evaluate the predictive accuracy of migration forecasts expressed in terms of probabilities. An attractive feature of the Brier score is that it can be decomposed into components of forecast accuracy that index both calibration and discrimination (Bessler and Ruffley, 2004, p. 400).

A comparative assessment of estimated models just focusing on the overall score can provide misleading information, as a superior score in absolute terms does not imply “sufficiency” or “unambiguous superiority” in all attributes of forecast accuracy (Murphy in Winkler, 1996, p. 48). Murphy (in Winkler, 1996, p. 44) emphasised that “decompositions provide quantitative measures of aspects of quality that are...
confounded when evaluation of forecasting performance is limited to the overall scoring rule”. Assessing probability forecasts based on a scoring rule and its attributes “at least implicitly recognizes that evaluation problems are multi-dimensional in nature and that forecast quality consists of several distinct attributes” (Murphy in Winkler, 1996, p. 44). The decomposition of Brier scores into specific attributes such as calibration and discrimination, the graphical depiction, and measures related to the joint distribution of forecasts and outcomes are much more valuable than the overall score (Winkler, 1996, p. 55).

The useful information provided by decompositions beyond the overall score suggests that for feedback and learning purposes, a scoring rule with convenient and well-understood decompositions should be selected (Winkler, 1996). Quadratic rules are most commonly used in scoring rules in practice, as they are easy to decompose, according to Winkler (1996). The Brier score has received considerable attention in meteorological literature (Murphy and Winkler, 1977; Winkler, 1996). It is a popular measure for ex-post evaluation of probability forecasts. Johnstone (2002) demonstrated that the Brier score is superior to categorical measures to accurately portray forecasting performance. However, the application of the Brier score in finance studies has been limited. Samuelson and Rosenthal (1986), Bessler and Ruffley (2004), and Yao et al. (2005) are among the few studies that have applied this scoring rule to assess the predictive performance of estimated models.

There are three approaches to decompose the Brier Score, as proposed by Sander (1963), Murphy (1973) and Yates (1982). The Sanders (1963) and Murphy (1973) decompositions require the forecasts to be sorted into a fixed set of categories, i.e. deciles. On the other hand, the covariance decomposition proposed by Yates (1982, p. 138) can be applied to probability forecasts that are either continuous or discrete. The
covariance decomposition provides components of forecast accuracy that are more basic than the Sander and Murphy decompositions (Yates, 1982, p. 141). The Yates partitions include calibration (bias), discrimination (slope) and a depiction of variability of forecasts (scatter), whereas Sander and Murphy decompositions consist of reliability (calibration) and resolution (discrimination) indices.

Furthermore, the components of the Yates decomposition can be displayed graphically and interpreted easily. The graphical depictions also provide useful statistics related to the joint distribution between forecasts and outcomes. Given its positive features, this thesis applies the Yates partitions to assess the predictive accuracy of migration forecasts. This method was used by Bessler and Ruffley (2004) to evaluate probability forecasts of stock market returns.

3.3. Conclusions

The review of literature on the discrete time cohort framework and the qualitative response static models, which are commonly used to estimate rating migration and default probabilities, led to the conclusion that these models had significant drawbacks in this application. Survival analysis was identified as an attractive alternative to the two foregoing approaches. Survival analysis has as its purpose the estimation of the probability of surviving in a given state over time. Thus it is well suited to estimating the probability that a rating grade will remain unchanged over time.

Cox’s hazard model (Cox, 1972) is the pre-eminent method to estimate survival models. It has been used in bankruptcy studies and some studies of rating migrations. Cox’s hazard model does not make any assumption about the distribution of migration times and it can be adapted to model competing risks. The Cox’s hazard model offers
the possibility to incorporate time-varying covariates, and to examine the changes of firm-specific and macro-economic risk factors on the hazard of credit events.

The dynamic Cox’s hazard model accommodates the evolution of risk factors over time. However, it has been a challenge to generate forecasts of time-varying survival probabilities in the presence of time-varying covariates. This has been due to computational issues in estimating the baseline hazard function. The issues become more complicated in a stratified dynamic Cox’s hazard model where multiple strata require multiple baseline hazard functions to be estimated.

Recent medical studies by Chen et al. (2005) applied an approach proposed by Andersen (1992) to estimate integrated baseline hazard function and overcome the computational issues. Utilising their technique, it is possible to generate time-varying survival forecasts in the presence of time-varying risk factors.

In regard to assessment of forecasting performance, three measures were discussed. The ROC and CAP measures, which are commonly used to evaluate the discrimination power of credit risk models, and the Brier score, which is widely used to assess probability forecasts in meteorology.

An attractive feature of the Brier score is that it can be decomposed into the components of forecast performance such as calibration and discrimination, while the ROC and CAP measures focus on discrimination. Of the methods available to decompose the Brier score, the covariance decomposition proposed by Yates (1982) is the most comprehensive and allows a convenient graphical interpretation.

The literature reviews above lay out the foundation for the method to be applied in this thesis. Building on this foundation, subsequent chapters apply the Brier score and
the Yates covariance decomposition to shed light on the forecast performance of the stratified Cox’s hazard models. The following questions are addressed:

(i) How accurate are the dynamic survival forecasts in a static estimation framework?

(ii) Does controlling for the evolution of macro-economic conditions improve forecast performance?

(iii) Does accounting for the interactions between time and rating history improve predictive performance?
Chapter 4

Method and data

The first part of this chapter presents details of the method used to estimate the migration models. The second part defines the firm-specific and macro-economic variables to be employed in the estimated models. The third part describes the data used in estimation of the models and in appraisal of the models’ forecast performance.

4.1. Method

4.1.1. Rating states

A rating state starts from the time a firm enters a rating class (start rating) subsequent to the commencement date of the observation period. The state ends at the time a firm either migrates to another rating class (end rating), becomes unrated (NR), or the observation period terminates. The time a firm keeps the same rating is the survival time. However, if a firm exits from a rating class due to a merger, extinction of rated debt, or any other reason apart from the event of interest (a downgrade or an upgrade), the survival time is treated as censored. Rating states commencing before the start of the observation period, or ending after the end of the observation period, are also treated as censored. The treatment of rating states in the estimation and holdout periods are depicted in Figure 4-1 below.

Survival analysis can utilise information pertaining to the time period during which left-censored observations are in the study (Yamaguchi, 1991; Blossfeld and Rohwer, 1995). It follows that the survival durations of left censored observations, state A and state B2, are measured from the beginning of the estimation and holdout period respectively.
Rating states like B1 (estimation period) and G (holdout period) are right-censored as they reach the termination date without experiencing the event. Consequently, the time of the event and the duration until the event occurs are not known. This type of right-censored observation introduces no bias into parameter estimates as the date of censoring is independent of the process that governs the occurrence of the event (Allison, 1995).

Rating state E is also right-censored as it becomes unrated (NR) during the estimation period. As discussed in chapter 3, becoming unrated might substitute for being downgraded, as issuers of deteriorating credit quality may choose to bypass credit rating agencies. If so, these unrated observations could represent informative censoring. Selectivity bias arises as these unrated observations “hide” information that is associated with the process of an event (Blossfeld and Rohwer, 1995). Sensitivity
analysis, as suggested by Allison (1995, p. 249-252), was conducted to investigate whether informative censoring introduces bias into the estimation process of the Cox’s proportional hazard model. It is found that the results of the estimated models developed in chapter 5 did not change substantively. Informative censoring, therefore, is not a concern to the empirical studies carried out in this thesis.

States E and C (estimation period) and state H (holdout period) are not censored, as they experienced the event during the respective observation period.

The duration of each rating state, until transition or censoring, was measured. The completed transitions were then labelled as upgrades, downgrades, or censored, according to their ending rating state. These rating transitions were then pooled across time, across issuers and across rating grades. As a result, a firm may contribute several rating transitions to the dataset.

The use of multiple observations for the same firm has the potential to introduce dependence among the observations (rating states). However, this problem is diminished to the extent that the covariates in the models control for dependence. In addition, this thesis uses the conditional gap time approach suggested by Prentice et al. (1981) to account for dependence among rating states of the same issuers. The alternative approaches proposed by Andersen et al. (1993), and Wei et al. (1989) assume that recurrent migrations of the same firm are independent and treat repeated migration events of different sequences alike. The conditional gap time approach, in contrast, takes the view that a rating is only at risk of a subsequent migration once prior migrations have occurred (Hosmer et al., 2008). This approach records the time sequence of repeated migration events, and rating data are stratified by the number of migrations the issuers have experienced since the observation period started.
There are multiple ending rating grades that a rating state could migrate to, that is, a rating state can be simultaneously at risk for multiple migration types. The approach taken in this thesis is to group together all the downgrades to the current state as down states, and all the upgrades to the current state as up states. Migration dynamics of the rating state are modelled as competing risks. As a result upgrades are treated as censored in the downgrades model, and downgrades are treated as censored in the upgrades model\textsuperscript{18}.

Attempting to estimate rating transitions between individual grades produces two problems. First, small sample sizes and sparse events result in low statistical power. This problem is intensified by the need to estimate stratified models and hence multiple baseline hazards. A focus on generic downgrade and upgrade events allows a larger number of states in the at-risk population. Second, it would be difficult to assess the model predictive performance for rare migration events. For example, the historical (actual) zero default frequencies in some high-end investment grades cannot be used as a reasonable benchmark to evaluate the predictive accuracy of the model. By pooling states across rating grades, this research can obtain sufficient historical transition data to be able to compare the observed migration rates with the probability estimates. Rating states must pass the screening test of having experienced at least two prior migrations. This ensures there is a rating history for each rating state studied. The stratified Cox’s hazard model uses both completed transitions and censored observations that pass the screening test. The estimation procedure makes use of risk sets, which are composed of all the firm ratings that are at risk of a rating change at time $t$. In the process of estimating the model a new risk set is formed at

\textsuperscript{18} Partington and Stevenson (2001), Yao, Partington, and Stevenson (2005) employed survival analysis framework to develop similar upstate and downstate Cox’s proportional hazard models of price reversal in the UK real estate market and Australian stock market respectively.
each time \( t \) when a rating transition occurs. Firm ratings leave the risk set once they experience a rating transition, or when they are censored. In forming the risk sets for the upgrade model, downgrades are treated as censored and vice versa. Upgrades and downgrades are thus treated as competing risks.

Rating states can be arranged in either a calendar time risk set or an event time (gap time) risk set framework. In the calendar time approach, durations are measured in calendar time. In this framework the macro-economic covariates are the same across all members of the risk set at each event time. The measurements of macro-economic covariates for rating states in a calendar time risk set are depicted in Figure 4-2.

**Figure 4-2**

*Measurements of macro-economic covariates for estimation states in the calendar time risk set*

- **Legend**
  - **Calendar time point** \( t \) at which a time-varying macro-economic variable is measured and its value is carried forward to represent the macro-economic condition prevailing at the subsequent start of rating state \( i \). For instance, the values of macro-economic covariate \( Z \) at the beginning of state E, D, C, B1 is \( Z^E(1), Z^D(0), Z^C(2), Z^{B1}(1) \) respectively.
  - **Calendar time point** \( t \) at which a time-varying macro-economic covariate is updated.
  - **The calendar time** \( t \) when a migration occurs and a new risk set is formed.

- **Estimation period**
- **Migrated state (Event)**
- **Censored state**

- **Time periods T1 and T2**
Macro-economic data are measured quarterly, and so the date of measurement will be
the same as the date of a migration event only by chance. Thus, when using macro-
economic data to form time-varying covariates, the values used in the estimation
process are updated to the most recent quarterly value as each risk set is formed.
Figure 4-2 above shows that the risk set formed at time $T_1$, $R(t=T_1)$, includes states B, C and D. The value of the time-varying macro-economic covariate $Z$ is equal to $Z(3)$ for all rating states in the risk set formed at time $T_1$, and equal to $Z(4)$ for all rating states C, B in the risk set formed at time $T_2$, $R(t=T_2)$. Since all the cases in a risk set have the same values for macro-economic variables at time $t$, the estimated model is “immune” to macro-economic influences. There is no variation across the risk set that allows the detection of any macro-economic effects.

In contrast, the event time (gap time) risk set arrangement “resets the clock” after a
migration event occurs, that is, time is measured from the last event (Hosmer et al.,
2008, p. 294). This arrangement is particularly relevant to the research carried out in
this thesis as the focus is on the duration between successive migrations rather than
the full course of recurrent events. Another positive feature of the event time (gap
time) approach is that it captures firm-specific macro-economic influences at each
event time. The arrangement of rating states in the event time risk set is depicted in
Figure 4-3 below.

In the gap time setting, the effect of macro-economic covariate $Z$ at event time $T_1$ for
rating state B is $Z(3)$, for state C is $Z(4)$, and for state D is $Z(2)$. The covariate at
event time $T_2$ includes $Z(4)$ for state B and $Z(3)$ for state D. The event time (gap
time) risk set framework therefore accommodates changes in the economic
environment over time and features firm-specific macro-economic effects at every
migration time.
Figure 4-3

Arrangement of rating states in the event time risk set

Legend

- The event time $t$ when a migration occurs and a new risk set is formed. The event times (or survival durations) of states C, D in the estimation period are respectively $T_1$, $T_2$.
- For $t=T_1$, the risk set formed at that time $R(T_1)$ includes state B1, C, and D.
- For $t=T_2$, the risk set formed at that time $R(T_2)$ includes states B1, D.

4.1.2. Model estimation

4.1.2.1. Estimation procedure

This study adopts the framework of event history analysis (Allison, 1984) and the Cox’s hazard model (Cox, 1972) to estimate the hazard of a downgrade or an upgrade, and to investigate the effect of rating history and time on rating migration hazards over the period 1984-2000.

The hazard rate is the rate of change of the survival probability over an interval, conditional on survival until the start of that interval. The hazard for a rating change of rating state $m$ is defined as follows:
\[
 h_m(t, Z) = h_{0}(t) \exp[Z^m_j \beta_j]
\]

Where \(h_m(t, Z)\) is the migration hazard of rating state \(m\) at time \(t\) given its time-fixed covariate vector \(Z^m_j\).

\(h_{0}(t)\) is the baseline hazard, which is the hazard with the covariate vector set to zero, at time \(t\). In an unstratified model, this baseline hazard would apply uniformly to rating states in the entire universe of the estimation period.

\(\beta_j\) is the vector of the estimated coefficients for the time-fixed covariate vector \(Z^m_j\).

The proportional hazard model in equation (1) applies the counting process framework and does not distinguish between recurrent events. This approach treats repeated migrations of the same firm uniformly and assumes recurrent events are independent. In addition, the model only incorporates the vector of time-fixed covariates \(Z^m_j\).

The stratified Cox’s hazard models developed in this thesis departs from the standard Cox’s proportional hazards model in various ways. First, the thesis applies the conditional gap time approach and develops the stratified Cox’s models to account for repeated migration events (Hosmer et al., 2008, pp. 208-211). Second, the thesis also develops stratified Cox’s hazard models with time-varying covariates to accommodate changes in the macro-economic environment (time-varying macro-economic covariates) and to account for the time spent in a rating state (time-varying interaction covariates) (Hosmer et al., 2008, pp. 213-216).

One stratified Cox’s proportional hazard model (time-fixed model), and two stratified time-varying covariate (TVC) hazard models were estimated for the upgrade and for the downgrade cases in the entire population. The time-fixed model incorporates all
time-independent covariates, such as rating history, industry classification, and time-fixed macro-economic covariates. The TVC base model extends the time-fixed model by including both time-fixed and time-varying macro-economic covariates. The TVC extended model adds further time-varying covariates that capture the interaction between rating history and time.

The stratified Cox’s proportional hazard model (stratified time-fixed model) may be expressed as follows:

$$h_{s,m}(t, Z) = h_{s(0)}(t) \exp[Z_j^m \beta_j]$$ \hfill (2)

Where $h_{s,m}(t, Z)$ is the migration hazard of rating state $m$ in stratum $s$ (migration sequence $s$) at time $t$ given its time-fixed covariate vector $Z_j^m$. $h_{s(0)}(t)$ is the baseline hazard for stratum $s$, which is the hazard with the covariate vector set to zero, at time $t$. This stratum-specific baseline hazard applies uniformly to all rating states in stratum $s$.

The stratified time-varying covariate hazard model can be expressed as follows:

$$h_{s,m}(t, Z, Z(t)) = h_{s(0)}(t) \exp[Z_j^m \beta_j + Z_p^m(t) \beta_p]$$ \hfill (3)

Where $h_{s,m}(t, Z, Z(t))$ is the migration hazard of rating state $m$ in stratum $s$ (migration sequence $s$) at time $t$ given its time-fixed covariate vector $Z_j^m$, and its time-varying covariate vector $Z_p^m(t)$.

$\beta_p$ is the vector of estimated coefficients for time-varying covariate vector $Z_p^m(t)$, and $\beta_j$ is the vector of estimated coefficients for time-fixed covariate vector $Z_j^m$.

In the TVC base model, $Z_p^m(t)$ contains time-varying macro-economic covariates. In
the TVC extended model $Z_p^m(t)$ includes both time-varying macroeconomic
covariates and covariates that capture the interaction between rating history and time.
In both TVC models, $Z^m_j$ contains rating history variables, industrial classification
dummies, and a time fixed macro-economic indicator.
The likelihood $L_{t,s}^m$ that rating state $m$ in stratum $s$ experiences a rating migration at
time $t_m$ is calculated as the state $m$’s hazard divided by the sum of the hazards of all
states in the risk set of stratum $s$ formed at event time $t_m$, $R(t_m, s)$.

$$L_{t,s}^m = \frac{\exp(\beta_j Z_j^m + \beta_p Z_p^m(t_m))}{\sum_{i \in R(t_m, s)} \exp(\beta_j Z_j^i + \beta_p Z_p^i(t_m))}$$

Where:
i represents a rating state in the risk set formed at time $t_m$ within stratum $s$, $R(t_m, s)$.

In equation (4), the values of time-varying covariates $Z_p(t_m)$ at the event time $t_m$ were
used for state $m$, which migrated at time $t_m$, and all observations $i$ within stratum $s$
that were in the risk set formed at time $t_m$, $R(s, t_m)$. The same state $i$ appearing in
different risk sets will carry different values of the time-varying covariates $Z_p(t)$
updated at various event times when those risk sets were formed.
To account for ties, in which several rating states experience the migration event at
the same time $t$, it has been traditional to use approximation adjustments such as the
Efron method (Efron, 1977). However, this study uses the exact method provided in
SAS version 9 for handling ties in the dataset19. This is the most precise method,
though it is computationally intensive, when there are a large number of ties.

Taking the product of the likelihoods, for all states that migrated, across all migration times $t_m$ observed in stratum $s$ gives the stratum $s$’s partial likelihood, $PL_s$, as follows:

$$PL_s = \prod_{m=1}^{n_s} L_{s,m}^{m} = \prod_{m=1}^{n_s} \left[ \frac{\exp \left( \beta_j Z_j^n + \beta_p Z_p^n(t_m) \right)}{\sum_{i \in R(t_m,s)} \exp \left( \beta_j Z'_j(t_m) + \beta_p Z'_p(t_m) \right)} \right]$$

(5)

Where:

- $t_m$ is the $m^{th}$ observed value of event times in the stratum $s$
- $m$ indexes a firm’s rating that experienced a migration at time $t_m$
- $n_s$ is the number of migration times observed in the stratum $s$

The partial likelihood for all risk sets (the full partial likelihood) is obtained by taking the product of the stratum’s partial likelihoods across all strata.

$$PL_{full\_stratified} = \prod_{s=1}^{N_{stratum}} PL_s = \prod_{s=1}^{N_{stratum}} \prod_{m=1}^{n_s} L_{s,m}^{m}$$

(6a)

$$PL_{full\_stratified} = \prod_{s=1}^{N_{stratum}} \prod_{m=1}^{n_s} \left[ \frac{\exp \left( \beta_j Z_j^n + \beta_p Z_p^n(t_m) \right)}{\sum_{i \in R(t_m,s)} \exp \left( \beta_j Z'_j(t_m) + \beta_p Z'_p(t_m) \right)} \right]$$

(6b)

Where

- $N_{stratum}$ is the number of strata observed in the entire population of rating states during the estimation period.

The vectors of the estimated coefficients $\hat{\beta}_p$ and $\hat{\beta}_j$ can be obtained by maximising the full partial likelihood in equation (6b).

The variance of a coefficient estimate in a stratified hazard model is a function of the total sample size and the total number of migration times (Hosmer et al., 2008). The
variance, therefore, is not a function of the stratum size or the number of events observed in any stratum.

4.1.2.2. Estimated baseline hazard function

It is evident from equation (5) that the baseline hazard is not used in the estimation process, as it cancels out in the numerator and denominator when forming the likelihood function. The baseline hazard is thus left unspecified, but of course it is required to estimate the probability of future rating transitions. This presents no problem when the model only contains time-fixed covariates. In that case the hazards of any two firms \( m \) and \( n \) in the risk set \( R(s, t_m) \) have a constant proportion through time as follow, and this property can be exploited to back out the baseline hazard.

\[
\frac{h_{s,m}(t, Z)}{h_{s,n}(t, Z)} = \frac{h_{s(0)}(t) \exp[Z_j^m \beta_j]}{h_{s(0)}(t) \exp[Z_j^n \beta_j]} = \exp[\beta_j(Z_j^m - Z_j^n)]
\]

However, with time-varying covariates the proportionality assumption does not hold, as can be seen from:

\[
\frac{h_{s,m}(t, Z, Z(t))}{h_{s,n}(t, Z, Z(t))} = \frac{h_{s(0)}(t) \exp[Z_j^m \beta_j + Z_p^m(t_m) \beta_p]}{h_{s(0)}(t) \exp[Z_j^n \beta_j + Z_p^n(t_m) \beta_p]} = \exp\{\beta_j(Z_j^m - Z_j^n) + \beta_p[Z_p^m(t_m) - Z_p^n(t_m)]\}
\]

It is not possible to readily extract the baseline hazard from the Cox’s regression results. Estimating the baseline hazard function \( \hat{h}_{s(0)}(t) \) in the presence of time-varying covariates is challenging.

Furthermore, while the unstratified hazard model as in equation (1) allows the estimation of one baseline hazard function that applies uniformly to all observations, the stratified time-varying covariate hazard model as in equation (3) requires the
estimations of multiple stratum-specific baseline hazard functions. Complicated issues arise, as the data set includes issuers of frequent rating changes and accordingly multiple strata are involved. Forming the estimation of multiple stratum-specific baseline hazard functions and generating time-varying probability forecasts for the rating data with numerous migration events is a daunting task.

This study applies a method proposed by Andersen (1992) and makes use of the SAS codes published by Chen et al. (2005) to estimate the stratum-specific integrated baseline hazard function. Given the vectors of the estimated coefficients \( \hat{\beta}_p \) and \( \hat{\beta}_j \) obtained from equation (6b), the integrated baseline hazard function \( H_{s(0)}(t) \) can be estimated as:

\[
\hat{H}_{s(0)}(t) = \sum_{t_m \leq t} \frac{D_m}{\sum_{i \in R(t_m, s)} \exp \left( \hat{\beta}_j Z_{ij} + \hat{\beta}_p Z_{ip}(t_m) \right)}
\]

Where

- \( D_m \) is an indicator of the event experienced by state \( m \) in stratum \( s \) at time \( t_m \) within the interval \([0,t]\).
- \( i \) represents a rating state in the risk set formed at event time \( t_m \) within stratum \( s \), \( R(t_m, s) \).

The estimation process requires the updated values of time-varying covariates \( Z_p(t) \) at each event time for all observations \( i \) in the risk set formed at that event time, (Andersen, 1992).

The stratum-specific integrated baseline hazard function \( H_{s(0)}(t) \) can also be expressed as a step function which is discontinuous at event time \( t_m \) (Chen et al., 2005).
\[ H_{s(0)}(t) = \sum_{t_m \leq t} [h_{s(0)}(t_m) (t - t_{m-1})] \]  

(8)

The estimated baseline hazard function \( \hat{h}_{s(0)}(t) \) that applies uniformly to all observations in stratum \( s \) can be derived from equations (7) and (8). The use of a step function is well suited to this study, as numerous events are observed and migration times are not far apart. The narrow gaps between successive event times should allow quite accurate estimation of the baseline hazard functions.

While the variance of a coefficient estimate \( \hat{\beta}_p \) or \( \hat{\beta}_j \) is a function of the total sample size and the total number of migration events, the variance of the estimated baseline hazard function \( \hat{h}_{s(0)}(t) \) depends on the number of observations and the number of event times observed in stratum \( s \) (Hosmer et al., 2008). In other words, the estimated baseline hazard function of a stratum with small sample size and few event times has more variance than the estimate of another stratum with many observations and numerous event times. In this study there are plenty of observations within strata 1 to 7 but with only 1.5 percent of the estimation sample spread over strata 8 to 14, accuracy of the baseline functions in these strata is a real concern.\(^{20}\) However, the maximum number of rating events which a holdout rating state experienced is 7. Therefore, the baseline functions in strata 8 to 14 are not used to generate survival forecasts in the holdout period.

4.1.2.3. Predicted survival function

The hazard function of state \( q \) in the holdout sample can be estimated using state \( q \)’s actual covariate vector \( Z_j^q \) and \( Z_p^q(t) \), the estimated baseline hazard function \( \hat{h}_{s(0)}(t) \)

\(^{20}\) The strata number indicates the number of recurrent events. Thus, stratum 7 contains issuers with seven rating migrations.
derived from equation (7) and (8), and the estimated coefficient vector $\hat{\beta}_p$ and $\hat{\beta}_j$ obtained from equation (6b).

The estimated hazard of holdout state $q$ in stratum $s$ can be calculated as:

$$\hat{h}_{s,q}(t, Z, Z(t)) = h_{s(0)}(t) \exp[Z^q_{s} \hat{\beta}_j + Z^q_{p}(t)\hat{\beta}_p]$$

(9)

The predicted survival function of a holdout state $q$ in stratum $s$ at time $t$ can be estimated as:

$$\hat{S}_{s,q}(t, Z, Z(t)) = \exp[-\sum \hat{h}_{s,q}(t, Z, Z(t))]$$

(10)

The survival profile for a holdout state $q$ can be derived from equation (10) by allowing $t$ to vary over the entire permissible range over which the time-varying baseline hazard function $h_{s(0)}(t)$ can be estimated and the time varying covariates $Z^q_{p}(t)$ for holdout state $q$ can be measured.

### 4.1.2.3.1. Forecast horizons

The choice of the appropriate forecast horizons $t$ depends on the purpose of forecasting rating migrations, and the forecast horizon should coincide with the risk horizon. For example, portfolio models may employ a one-year forecast horizon to calculate credit risk exposures, whereas credit pricing models may require shorter forecast horizons. Bank loans or seasoned bonds may require one-year and five-year horizons respectively, while a ten-year forecast horizon is appropriate for newly issued bonds or private placements (Altman, 1998). The conventional forecast horizon is one year, and, in practice, credit rating agencies publish annual transition matrices. Since credit monitoring generally follows a quarterly review pattern and macro-economic time series are published on a quarterly basis, short-term forecasts were
constructed at quarterly intervals within a one-year window ($t = 0.25, 0.5, 0.75, 1$ year). These forecasts are suitable for issuers with a volatile rating history, and capture the stylised fact that issuers on their downward journey only stay a short time in the middle-rating grades (Lando and Skodeberg, 2002; Koopman et al., 2006).

As the subjects of the study are generally seasoned issuers, longer-term forecasts were also formed at yearly intervals within a five-year window ($t = 1, 2, 3, 4, 5$ years). These forecasts are appropriate for issuers of more stable rating history.

In forming these forecasts the approach of Chen et al. (2005) is followed. Chen et al. (2005) assumed that all observations in the holdout sample survive at the shortest forecast horizon. In this study, the shortest horizon is one quarter (short-term forecasts) and one year (longer-term forecasts). As the horizon unfolds, Chen et al. (2005) deleted from the holdout sample at time $t$ those cases which were censored, or had experienced the event, before time $t$. Censored cases have to be deleted since their status is not known at time $t$. Deleting the cases that have experienced the event means that the holdout sample reflects the changing population, but has the disadvantage that ex-post information is used in forming the holdout sample. This would be expected to bias forecast performance upwards; however, there is an offsetting effect. The form of the migration models is such that the estimated probability of migration increases with the passage of time. Keeping firms that have already migrated in the holdout sample rather than deleting therefore leads to an apparent improvement in the accuracy of the model as time passes.

The approach of Chen et al. (2005) results in a holdout sample that gets smaller with the passage of time. Consequently variations in the forecast performance of the model
through time are a consequence of both the passage of time and the changing sample composition.

The time-fixed model and the TVC base model generate unconditional survival forecasts. They are unconditional in the sense that the models do not capture the increasing duration as rating states continue unchanged (the interaction terms between rating history and time are not included in the models). Table 4-1 summarises the time horizons at which survival probabilities are estimated by the time-fixed and the TVC base models for each holdout state:

<p>| Table 4-1 |
| UNCONDITIONAL FORECAST HORIZONS |</p>
<table>
<thead>
<tr>
<th>Forecast horizons t (year)</th>
<th>Short-term</th>
<th>Longer-term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional hazard model (with time-fixed firm-specific and macro-economic covariates)</td>
<td>0.25 0.5 0.75 1</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>TVC base model (with time-fixed firm-specific, time-fixed and time-varying macro-economic covariates)</td>
<td>0.25 0.5 0.75 1</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>

4.1.2.3.2. The construction of macro-economic covariates and time interaction terms for holdout states

The values of macro-economic covariates $Z_p(t)$ used in the estimation process (equation 3 and 4) are updated to the most recent quarterly lagged value as each risk set is formed (see Figure 4-2 and 4-3).

A forecaster does not know, at the beginning of holdout states, the times of subsequent rating changes and future macro-economic developments. Only information up to the commencement of holdout states (time zero) is known. In other words, at time zero it is impossible to get quarterly updated macro-economic values over the survival durations of holdout states. As shown in Figure 4-4 below, the value
of the macro-economic covariate $Z_p^q(t)$ used in equation (9) to form forecasts for holdout state $q$ was measured at state $q$’s commencement (calendar time $t=5$ for state G, $t=6$ for state H, and $t=5$ for state B, or event time $t=0$ for all states), and entered equation (9) without being subsequently updated.

**Figure 4-4**

The construction of macro-economic covariates for holdout states

---

Calendar time point $t$ at which a macro-economic variable is measured and its value is carried forward to represent the macro-economic condition prevailing at the subsequent start of holdout state $q$

Calendar time point $t$ at which a macro-economic covariate is measured

---

Figure 4-5 below depicts the constructions of time interaction covariates for observations in the estimation and holdout samples. For the estimation process, the interaction term $Z_p^i(t)$ in equation (3) and (4) is constructed as a function of the rating history of estimation state $i$ and the event time $t$, and is updated whenever an event occurs ($t=T_1, T_2$). For instance, the time interaction covariate $Z(t)$ at $t=T_1$ includes $Z_B^{1}(T_1), Z_C^{1}(T_1), Z_D^{1}(T_1)$, in which $Z_B^{1}(T_1)$ is a function of the rating history of state B1 and the event time $T_1$, $Z_C^{1}(T_1)$ is a function of the rating history of state C and event time $T_1$, and so on. The time interaction covariate $Z(t)$ at $t=T_2$ includes $Z_B^{1}(T_2), Z_D^{1}(T_2)$. 

81
The construction of the interaction terms $Z_p^n(t)$ between rating history and time for holdout state $q$ is more complicated. As the exact times of rating changes for holdout states are not known at time zero, the interactions term $Z_p^n(t)$ for holdout state $q$ cannot be constructed in the same way as $Z_p^i(t)$ for estimation state $i$. Instead, the time interaction covariates $Z_p^n(t)$ used in Equation (9) to form forecasts for holdout state $q$ at time $t$ are constructed conditional on the holdout state $q$ not migrating and not having its rating withdrawn at time $T^*$ with $T^*<t$. The values of a time interaction covariate $Z_p^n(t)$ in equation (9) are taken as a time-varying function of the rating history variable of state $q$ and the conditional survival time $T^*$, $Z_p^n(t = T^*)$. $T^*$ can take different values over the forecast horizon $t$. 
With the inclusion of the time-varying interaction term $Z_p^q(t = T^*)$ in equation (9), the TVC extended models generate forecasts that are conditional on the length of time ($T^*$) a holdout state survives in its current rating grade. While unconditional forecasts differ in forecast horizons and forecast intervals, conditional forecasts differ not only in horizons and forecast intervals but also in the conditional survival durations of holdout states. Two generic types of conditional forecasts are formed for each holdout state:

Short-term forecasts are formed at quarterly intervals conditional on holdout states surviving at one-quarter lead time ($T^* = 0.25$ year). That is, forecasts at two-quarter ($t = 0.5$ year), three-quarter ($t = 0.75$ year), and one-year horizons ($t = 1$ year) are formed conditional on holdout states surviving at one-quarter lead time ($T^* = 0.25$ year).

Longer-term forecasts are generated at yearly intervals conditional on holdout states surviving a progressively longer period of time and not experiencing the event at one-year ($T^* = 1$ year), two-year ($T^* = 2$ years), three-year ($T^* = 3$ years), and four-year ($T^* = 4$ years) lead time. For instance, forecasts at two-year ($t = 2$ years), three-year ($t = 3$ years), four-year ($t = 4$ years), and five-year ($t = 5$ years) horizons are generated conditional on holdout states surviving at one-year lead time ($T^* = 1$ year). Forecasts at five-year horizon ($t = 5$ years) are formed conditional on holdout states surviving at four-year lead time ($T^* = 4$ years), and so on.

Table 4-2 summarises, for the TVC extended model, the dates the rating is assumed to survive to (conditional survival time $T^*$) and the horizons $t$ over which the probability of survival is then forecasted.
Given the evidence provided by Kiefer and Larson (2007) and Frydman and Schuermann (2008), that within one or two years the Markov property adequately holds, it is appealing to examine the predictive performance of rating history over the one-year and two-year horizons. Furthermore, a sufficiently large number of holdout observations are available at these horizons.

4.1.2.4. Forecast evaluation

4.1.2.4.1 Brier score

This study uses a proper scoring rule, the Brier score (Brier, 1950), to assess the predictive accuracy of the estimated models for a holdout sample formed during the period 2001-2005. The actual survival status of each holdout state is recorded, and mapped against the survival probability estimate obtained at the forecast time $t$. The issuer with the lowest survival probability estimate is the most risky with respect to the migration event relevant to each model. For instance, in the upgrade model the firm with the lowest survival estimate is the one most at “risk” of being upgraded.

The Brier score at forecast time $t$, $B_t$, is defined as follows:

$$B_t = \frac{\sum_{q=1}^{N_t} (f_{t, \text{state}_q} - a_{\text{state}_q})^2}{N_t}$$  \hspace{1cm} (11)$$

Where
$f_i^{state-q}$ indicates the probability forecast that the holdout state $q$ will survive at forecast time $t$.  

$a^{state-q}$ is the known outcome survival state of holdout state $q$. If holdout state $q$ survives, $a^{state-q} = 1$, and if holdout state $q$ experienced the migration event of interest (i.e. a downgrade in the downgrade model or an upgrade in the upgrade model), $a^{state-i} = 0$.

$N_t$ is the number of forecasts at forecast time $t$. As the forecast horizon expands, the number of forecasts decreases.

The Brier score varies from 0 to 1. A Brier score of zero indicates perfect prediction ability and a Brier score of 1 indicates the worst prediction ability. A lower Brier score therefore implies better forecast performance.

The Brier score at forecast time $t$ will be assessed by reference to a naïve Brier score and a benchmark Brier score. The naïve Brier score takes the value of 0.25 and is formed by setting the predicted survival probability $f_i^{state-q}$ in equation (11) equal to a random forecast of 0.5. The benchmark Brier score is obtained from equation (11) by setting the predicted survival probability $f_i^{state-q}$ equal to the proportion of states that actually survived beyond time $t$ in the estimation sample.

### 4.1.2.4.2. Brier score decomposition

Unlike traditional measures common in the evaluation of the discriminatory power of credit risk models such as ROC and CAP curves, the Brier score can be conveniently decomposed into components of forecast accuracy such as calibration (reliability) and

---

22 The notation was changed to provide a compact presentation of the formula in a form consistent with the literature review on the Brier score.
discrimination (resolution). This thesis applies the covariance decomposition proposed by Yates (1982, pp. 138-141) to decompose Brier scores and assess the prediction ability of estimated models over different forecast horizons.

The most basic form of the covariance decomposition of Brier score $B_t$ at the forecast time $t$ is given as:

$$B_t = \bar{d}_t (1-\bar{d}_t) + (\bar{f}_t - \bar{d}_t)^2 + S_{\bar{f}_t}^2 - 2 \bar{f}_t \bar{d}_t$$  \hspace{1cm} (12)

Uncertainty    Bias square   Scatter   Covariance

Where:

$\bar{d}_t$, or $\bar{d}$ for short, is the overall mean survival index, or the survival base rate at time $t$

$\bar{d}_t (1-\bar{d}_t)$, or $\bar{d}(1-\bar{d})$ for short, is the variance of the outcome index at the forecast time $t$

$\bar{f}_t$, or $\bar{f}$ for short, is the overall mean survival forecast at the forecast time $t$

$(\bar{f}_t - \bar{d}_t)$, or $(\bar{f} - \bar{d})$ for short, is the bias of forecasts at the forecast time $t$

$S_{\bar{f}_t}^2$, or $S_{\bar{f}}^2$ for short, is the variance of the forecasts, or scatter, at the forecast time $t$.

$S_{\bar{f}_t \bar{d}_t}$, or $S_{\bar{f} \bar{d}}$ for short, is the covariance of the survival outcome index $a_{\text{state} \rightarrow q}$ and the survival probability forecast $f_{\text{state} \rightarrow q}$ at the forecast time $t$. The survival outcome index $a_{\text{state} \rightarrow q}$ can take the value of 1 (if state $q$ survives) or 0 (if state $q$ migrates). The survival forecast $f_{\text{state} \rightarrow q}$ can take any value between 0 and 1, the higher the probability forecast, the more likely holdout state $q$ survive in its current rating grade.
The first component $\tilde{d}(1-\tilde{d})$, namely the outcome index variance, is calculated as the survival base rate $\tilde{d}$ multiplied by its complement $(1-\tilde{d})$. The survival base rate $\tilde{d}$ is determined by “natural forces”, and as a result the outcome index variance $\tilde{d}(1-\tilde{d})$ reflects an aspect of forecast accuracy that is not controlled by, and does not depend on the predictive skill of the forecaster (Yates, 1982, p. 139). Removing this component from the overall score “levels the playing field” and improves the validity of comparative assessments across different predictive models over the same sample at the same forecast time. The remaining components of the Brier score, namely bias, scatter and covariance, reflect the predictive ability of rating history in forecasting future migrations. These skill components are used in assessing the relative performance of estimated hazard models.

The first skill component, bias, is defined as the difference between the mean survival probability forecasts $\bar{f}$ and the mean survival outcome index $\tilde{d}$, or $(\bar{f} - \tilde{d})$. This term can be either positive or negative. The smaller the absolute value of bias, the lower the Brier score, the aim of the forecasters being to minimise bias. The square of bias is “reliability-in-the-large” or “calibration-in-the-large”. Bias indicates the ability of a forecaster to match the overall mean forecast $\bar{f}$ to the survival base rate $\tilde{d}$ (Yates, 1982, p. 140). Bias reflects the overall pessimism or optimism of the forecaster in assigning probability survival forecasts to holdout states.

The second skill component, scatter or $S^2$, is the pooled variance of the forecasts. This term is derived from the distribution of probability forecasts assigned to surviving states $f_1$ and forecasts assigned to non-surviving states $f_0$ (Arkes, Dawson, Speroff, Harrell, Alzola, Phillips, Desbiens, Oye, Knaus, Connors and the Support Investigators, 1995, p. 121). Scatter is defined as:
\[ S_f^2 = \frac{N_1 \text{Variance}(f_1) + N_0 \text{Variance}(f_0)}{N_1 + N_0} \]  \hspace{1cm} (13)

Where:

- \( N_1 \) is the number of surviving rating states
- \( N_0 \) is the number of non-surviving rating states

Scatter represents the “noisiness” of survival probability forecasts, and reflects the sensitiveness of the forecast to information that is not related to the survival of holdout states. The smaller the scatter, the lower the Brier score, the aim of the forecaster being to minimise scatter. The scatter, or variance of forecasts, \( S_f^2 \), takes a minimum value of zero when the model produces constant forecasts.

The third term of the skill components, covariance \( S_{fd} \), can be expressed as:

\[ S_{fd} = (\bar{f}_1 - \bar{f}_0)[\bar{d}(1-\bar{d})] \]  \hspace{1cm} (14)

Where

- \( \bar{f}_1 \) is the mean survival forecasts assigned to rating states that actually survive
- \( \bar{f}_0 \) is the mean survival forecasts assigned to non-surviving states, i.e. states that experience the event of interest.

Since the outcome index variance \( \bar{d}(1-\bar{d}) \) is out of the control of the forecaster, covariance is determined by the term \( (\bar{f}_1 - \bar{f}_0) \), which is the difference between average survival probability forecasts assigned to holdout states that survive and to those that do not survive (Yates, 1982, p. 138). In the covariance graph (discussed below), this term is the slope of the line that connects \( \bar{f}_1 \) and \( \bar{f}_0 \). Given a base rate \( \bar{d} \), the larger the term \( (\bar{f}_1 - \bar{f}_0) \) or the steeper the slope, the lower the Brier score. The
slope indicates the ability of the forecaster to distinguish between the group of surviving states \( f_1 \) and the group of non-surviving states \( f_0 \). A steeper slope reflects the forecaster’s ability to assign higher forecast probabilities to surviving states than to non-surviving states.

4.1.2.4.3. Covariance graph

Yates (1982), and Arkes, Dawson, Speroff, Harrell, Alzola, Phillips, Desbiens, Oye, Knaus, Connors and the Support Investigators (1995) showed that the covariance components of a Brier score can be easily graphed and depicted in a covariance graph. An example of the covariance graph, as described by Yates (1982, pp. 143-148) and Arkes et al. (1995, pp. 121-123), is presented in Figure 4-6 below. Probability survival forecasts were generated by the Cox’s proportional hazard upgrade model at a three-year horizon.

Figure 4-6

Brier score 0.1704, Scatter 0.0227, Bias -0.2064, Slope 0.0714
The abscissa shows the survival outcome index. The two possible outcomes for holdout states in the upgrade model are migration (upgrade), which is denoted as 0 on the left, and survival, which is denoted as 1 on the right. Of 273 holdout states available at a three-year forecast horizon, 234 states survive, and 39 states migrate (upgrade). A vertical dotted line (in pink) is located at the survival base rate, or the overall mean survival outcome index $\bar{d} = 0.8571$, on the abscissa. On the ordinate is the survival probability estimates categorised in deciles. A horizontal dotted line (in green) is located at the overall mean survival forecasts $\bar{f} = 0.6506$ on the ordinate.

The 45-degree solid line (in black) represents unbiased estimates. Bias can be measured as the vertical distance from the 45-degree line to the point where the vertical pink dotted line (survival base rate line) and the horizontal green dotted line (mean survival forecast line) cross. This point is denoted as a quadrilateral ($◊$). If a model produces unbiased forecasts, then the two dotted lines will cross on the 45-degree line, corresponding to a zero bias (Arkes et al., 1995). If the two dotted lines meet below the diagonal line, the model produces “pessimistic” forecasts. That is, the model underestimates the survival outcome, corresponding to a negative bias. In the covariance graph above, the model is about 20.64 percent too pessimistic in its three-year forecasts. On the other hand, should the two dotted lines cross above the diagonal, the model overestimates the survival outcome by making “optimistic” forecasts. In that case, the model has positive bias.

On the vertical lines above the survival outcome (1) and the migration outcome (0) indices are the histograms for survival forecasts of 234 states that actually survive and 39 states that are upgraded respectively. Surviving and non-surviving holdout states are stratified into distinct decile categories (bins) in order of estimated survival probabilities. In this setting, states with survival forecasts varying from 0 to 10 percent
are put together, those with forecasts ranging from 11 percent to 20 percent in another bin (decile) and so on. The bars on these histograms illustrate the percentage of survival forecasts made at the individual probability deciles. The number of survival forecasts observed within each decile was attached to the corresponding bar for ease of reference. Each histogram bar was centred at the average survival forecasts observed within the corresponding probability decile. The further the histogram bars spread along the vertical lines, the greater the scatter of the survival forecasts (Arkes et al., 1995). Though the magnitude of the scatter cannot be quantified, the histogram displays the distribution of survival forecasts across probability deciles and provides a rough description of the scatter.

The outcome index line extending vertically from 1 (on the right edge) includes a point denoted as a quadrilateral (◊). This represents the mean survival forecasts given to states that actually survive or $\bar{f}_1=0.6608$. Similarly, the outcome index line drawn vertically from 0 (on the left edge) contains another quadrilateral (◊). This denotes $\bar{f}_0=0.5895$, the average survival forecasts given to states that actually migrate (upgrade). The dotted line (in brown) linking these two quadrilaterals is the regression line for a survival forecast on the outcome index. The slope of the regression line is calculated as the difference between the average survival forecasts given to states that actually survive and states that actually migrate, or $(\bar{f}_1 - \bar{f}_0)=7.14$ percent. The regression line will be flat (slope=0) if the mean survival forecasts for states that survive and states that migrate (upgrade) are the same. The further the regression line diverges from the horizontal, the more discrimination in the forecasts of the two groups (Arkes et al., 1995). The regression line, the vertical base rate line, and the horizontal mean forecast line intersect at the point denoted as a quadrilateral (◊).
As shown above, the covariance graph sheds light on the forecast performance of an estimated model at a point in time. It displays basic accuracy measures such as outcome index variance, bias, scatter, and slope. In addition, it also provides statistics related to the distribution of probability survival forecasts across decile categories.

4.2. Measurements and definitions of variables

4.2.1. Measurements of rating history and macro-economic variables

4.2.1.1. Rating history numeric codes

The data on US corporate obligor rating history were supplied by Standard & Poor’s CreditPro2005 issuer dataset, and the study covers the period 1984 to 2005. Standard and Poor’s rating scale include 10 major rating categories varying from excellent credit quality to default as follows, AAA, AA, A, BBB, BB, B, CCC, CC, C, D. “The ratings from ‘AA’ to ‘CCC’ may be modified by the addition of a plus (+) or a minus (-) sign to show relative standing within the major rating categories.”23 Ratings above BB+ are investment rated whereas rating below BBB- are speculative (junk) rated.

Changes in ratings outlooks or CreditWatch listings of issuers have not been incorporated in developing the models. The reason for this is twofold. First, the importance of Outlooks or CreditWatch listings has been recently recognised in the literature, for example Hamilton and Cantor (2004), Vazza, Leung, Alsati, and Katz (2005), Al-Sakka and Gwilym (2009), Hill, Brooks and Faff (2010). Second, the Standard & Poor’s CreditPro 2005 dataset available at the University of Sydney did not contain Outlook and CreditWatch data24.

24 This data set was purchased by the Faculty of Economics and Business at the University of Sydney.
While previous studies on rating dynamics\textsuperscript{25} tend to focus on coarser rating categories such as A, B and so on, or investment and speculative grades, and neglect fine revisions intra-rating such as + and - distinctions, this thesis employs the full rating sub-categories such as AA, AA-, AA+. The use of a fine partition of rating grades increases the precision with which the rating history variables are measured.

In this thesis, the rating scales are coded from 0 to 26 with 0 indicating the default state (D) and 26 indicating the AAA grade. Rating changes are recorded by notches based on a numerical rating scale as in Table 4-3A below.

\textbf{Table 4-3A}

<table>
<thead>
<tr>
<th>Standard &amp; Poor’s rating</th>
<th>Numeric code</th>
<th>Speculative (Junk) grade</th>
<th>Speculative (Junk) grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>26</td>
<td>BB+</td>
<td>15</td>
</tr>
<tr>
<td>AA+</td>
<td>24</td>
<td>BB</td>
<td>14</td>
</tr>
<tr>
<td>AA</td>
<td>23</td>
<td>BB-</td>
<td>13</td>
</tr>
<tr>
<td>AA-</td>
<td>22</td>
<td>B+</td>
<td>12</td>
</tr>
<tr>
<td>A+</td>
<td>21</td>
<td>B</td>
<td>11</td>
</tr>
<tr>
<td>A</td>
<td>20</td>
<td>B-</td>
<td>10</td>
</tr>
<tr>
<td>A-</td>
<td>19</td>
<td>CCC+</td>
<td>9</td>
</tr>
<tr>
<td>BBB+</td>
<td>18</td>
<td>CCC</td>
<td>8</td>
</tr>
<tr>
<td>BBB</td>
<td>17</td>
<td>CCC-</td>
<td>7</td>
</tr>
<tr>
<td>BBB-</td>
<td>16</td>
<td>CC</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D (default)</td>
<td>0</td>
</tr>
</tbody>
</table>

The numeric conversion maintains the rank order of the rating scale, and assumes that the difference between any two consecutive rating states is the same. For instance, it is assumed that the “rating gap” between AA+ (24) and AA (23) is the same as the “rating gap” between CCC- (7) and CCC (8). The use of numeric scales to replace alphabetical rating scales was employed by Sy (2002), Kim and Wu (2008), Al-Sakka and Gwilym (2009), Hill, Brooks and Faff (2010).

The alternative of coding each rating class through dummy variables would consume a substantial number of degrees of freedom. Adding an extra 17 dummy variables to the model would also preclude compact presentation of the results and make interpretation very difficult, particularly in the case of the TVC extended models where interaction terms are introduced\textsuperscript{26}. In this thesis, two dummy variables indicating whether the current rating is in the boundary of investment (BBB-, BBB, BBB+) or speculative rating grades (BB-, BB, BB+) were included in the models. These dummies would help capture any non-linearity of the investment/ speculative boundary ratings relative to the neighbouring ratings.

To achieve the interval and ratio levels of measurement, necessary for other than rank order statistics, it is required that that the gap between scores, on the numerical transformation of the rating scale, be equal. An issue that arises here is the treatment of the rating gap between AAA and AA+ and the gaps between adjacent ratings below CCC-. The issue arises because for the rating grades above AA+ and below CCC-, the plus and minus notches are not employed by Standard and Poor’s. The question is whether the omission of notches leads to different gap lengths in the rating scale. In this thesis it has been assumed that the gap between AAA and AA+ is not one notch but two, implicitly including an AAA- notch in the rating. Thus the AAA grade is recoded to 26 and the AA+ grade is recoded to 24. A similar approach has been applied to

\textsuperscript{26} Robustness tests in which sub-rating categories were grouped into major rating categories and each major rating category was coded through a dummy variable were applied to the Cox’s proportional downgrade and upgrade hazard models. This thesis found that alternative coding major rating categories through dummy variables does not substantially change the results of the estimation models.
ratings below CCC-. As a consequence, the range of numerical scores is greater than the number of notches that Standard and Poor’s actually employ.\footnote{27}

The assumption that grades above AA+ and below CCC- cover multiple notches, is unlikely to make a substantive difference to the analysis, as with one exception only small numbers of migrations involve these grades. This is evident from Table 4-3B which summarizes the occurrences of migrations in the estimation period, for ratings to which multiple notches were imputed. The exception was migration to default where there were quite a few cases. However, given the substantial economic difference between being a defaulter and a non-defaulter a case can be made here for a multiple notch difference in scores.

Table 4-3B

<table>
<thead>
<tr>
<th>Migrations either departed from or ended at top and bottom rating grades</th>
<th>Of 1698 downgrades (including default)</th>
<th>Of 1175 upgrades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migrations from C to other rating grades</td>
<td>Number of defaults</td>
<td>Percentile of total</td>
</tr>
<tr>
<td>Migrations from CC to other rating grades</td>
<td>59</td>
<td>3.47%</td>
</tr>
<tr>
<td>Migrations from other rating grades to C</td>
<td>236</td>
<td>13.90%</td>
</tr>
<tr>
<td>Migrations from other rating grades to CC</td>
<td>45***</td>
<td>2.65%</td>
</tr>
<tr>
<td>Migrations from other rating grades to D (default)</td>
<td>236</td>
<td>13.90%</td>
</tr>
<tr>
<td>Migrations from AAA to other rating grades</td>
<td>6***</td>
<td>0.51%</td>
</tr>
<tr>
<td>Migrations from other rating grades to AAA</td>
<td>6***</td>
<td>0.51%</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
<td>3.71%</td>
</tr>
</tbody>
</table>

Note

* Downgrade from CC: 1 to C
** Downgrades to C: 2 from B, 1 from CCC+, 1 from CCC
*** Downgrades to CC: 1 from BB-, 1 from B, 10 from B-, 14 from CCC+, 14 from CCC, 5 from CCC-
^ Downgrades from AAA: 2 to AA+, and 1 to AA
^^ Upgrade from C: 1 to CC
^^^^ Upgrades to AAA: 1 from A-, 3 from AA, 2 from AA+

4.2.1.2. Macro-economic lag series

Whether the economy is doing well or badly will influence the risk of a rating migration. Therefore macro-economic variables have been used in the analysis to

\footnote{27 I acknowledge and thank Owain ap Gwillym for pointing out the difference between the numerical scores and the Standard and Poor’s ratings.}
control for the economic conditions prevailing at the beginning of each rating state (time-fixed covariate models) and prior to each event time (time-varying covariates, TVC models).

Seventeen potential macro-economic covariates were considered and those that exhibited strong multicollinearity were eliminated. Notable deletions were the credit spread and the default rate. This left seven covariates that were used in the hazard models, of which one covariate is time-fixed (recession indicator). The other six covariates are constructed as time-fixed in the static hazard models, and time-varying in the TVC base and TVC extended hazard models.

The recession indicator (dummy recession), the Chicago Fed National Activity Index (CFNAI), the output growth gap (RealGDPg actual minus potential), and the growth in industrial production level (Industrial production change) capture the general level of economic activity. The S&P 500 quarterly return (SP500 quarterly return), and S&P 500 annualised standard deviation (SP500 annual SD) represent the performance of the stock market, while the term structure slope (term structure slope) reflects credit condition.

As macro-economic conditions are likely to affect rating migrations with a lag, this thesis uses an exponentially weighted average of lagged observations computed quarterly to construct five of the time-varying macroeconomic covariates. The construction of macro-economic lagged values is similar to the approach applied by Figlewski et al. (2008). The exponentially weighted average value $X_t$ for the quarterly series $x$ for a given macro-economic variable in quarter $t$ is calculated using data up to the previous quarter as:
\[ X_t = \frac{\sum_{k=1}^{K} \delta^{k-1} x_{t-k}}{\sum_{k=1}^{K} \delta^{k-1}} \] 

(15)

Where

\( K \) is the length of the lagged window and \( \delta \) is the decay factor.

The lagged length of six quarters (\( K=6 \)) and the decay factor \( \delta = 0.6815 \) are used. These parameters are derived from Figlewski et al. (2008) by transforming their monthly weighting scheme (\( K=18 \) months and \( \delta = 0.88 \)) to an equivalent quarterly weighting scheme.

There was one exception in regard to the construction of lag series, the monthly Chicago Fed National Activity Index (\( CFNAI \)) is published as a three-month moving average, and this thesis uses the \( CFNAI \) without further transformation.

### 4.2.2. Definitions of variables

The Cox’s proportional hazard and TVC base hazard models contain 34 variables. Both models include 16 time-fixed rating history covariates, 11 time-fixed industry sector dummies and one time-fixed macro-economic covariate (dummy recession). The other six macro-economic covariates are constructed as time-fixed in the Cox’s proportional hazard model, and time-varying in the TVC (base and extended) hazard models. Definitions of the variables in the hazard models are given below.

#### 4.2.2.1. Rating history variables

*Lag one:* The duration (in years) of the rating state that ends with either a downgrade or an upgrade (ie. is not censored) and immediately precedes the current rating.
*Lag two:* The duration (in years) of the rating state that ends with either a downgrade or an upgrade, and immediately precedes the lag one rating.

*Dummy lag1 down:* This dummy captures the direction of the lag-one regrade and takes the value of one if the lag one rating ends with a downgrade, and zero otherwise.

*Dummy lag2 down:* This variable captures the direction of the lag-two regrade and takes the value of one if the lag-two rating ends with a downgrade, and zero otherwise.

*Rate prior change:* This indicates the average number of rating changes per year over the firm’s rating history. It is calculated as the number of prior migrations (downgrades and upgrades) observed between the entry of the firm to the study and the beginning of the current rating state divided by the period from the time of entry till the start of the current rating state.

*Rate prior down:* This equals the average number of downgrades per year over the firm’s rating history. It is calculated as the number of downgrades observed between the entry of the firm to the study and the beginning of the current state divided by the period from the time of entry to the study until the start of the current rating state.

*Age since first rated:* The rating age of the firm, which is equal to the length in years from the time the firm was first rated until the beginning of the current state.

*Original rating:* The rating of the firm when it was first rated.

*Start rating:* The rating at the beginning of the current rating state.

*Dummy inv boundary:* The dummy takes the value of one if the start rating falls within the investment (inv) boundary BBB-, BBB, BBB+, and zero otherwise.
**Dummy junk boundary:** This dummy takes the value of one if the start rating falls within the speculative (junk) boundary BB-, BB, BB+, and zero otherwise.

**Number NR (not rated):** This indicates the number of times a firm underwent a break in rating history from the entry of the firm to the study until the beginning of the current rating state$^{28}$.

**Number FA:** This indicates the number of times a firm experienced a fallen angel (FA) event (a downgrade from investment-grade state to speculative-grade state) from the entry of the firm to the study until the beginning of the current rating state.

**Number RS:** This indicates the number of times a firm experienced a rising star (RS) event (an upgrade from speculative-grade state to investment-grade state) from the entry of the firm to the study until the beginning of the current rating state.

**Number big down:** This indicates the number of times a firm experienced a big downgrade jump, defined as a jump of at least three numeric rating notches$^{29}$, from the entry of the firm to the study until the beginning of the current rating state.

**Number big up:** This variable indicates the number of times a firm experienced a big upgrade jump, defined as a jump of at least two numeric rating notches$^{30}$, from the entry of the firm to the study until the beginning of the current rating state.

---

$^{28}$ Being not rated is infrequently observed. No firm experienced more than two periods of being unrated and they represented less than 0.5 percent of the estimation sample. Those firms which experienced one period of being unrated account for only 6.73 percent of the estimation sample.

$^{29}$ Firms that experienced a downgrade of at least three rating notches account for 11.48 percent of the estimation sample and 17 percent of the holdout sample.

$^{30}$ Firms that experienced an upgrade of at least two rating notches account for 6.55 percent of the estimation sample and 4.8 percent of the holdout sample.
4.2.2.2. Control variables

*Industry dummies*: Firms’ industry sectors, as given by Standard & Poor’s, were used to control for any industry effects. The industry dummy took a value of one if the firm was in an industry sector and zero otherwise. Due to their unique business nature and credit risk exposure, firms in the financial institution sector\(^\text{31}\) were excluded from the study, which left 12 sectors in the study. The 12 industry sectors resulted in 11 dummy variables with the insurance sector left uncoded.

*Dummy recession*: This indicator is used to control for the macro-economic environment prevailing at the commencement of rating states. The dummy variable takes the value of one if the rating state started at the time of a recession, defined by the National Bureau of Economic Research (NBER)\(^\text{32}\), as 1 August 1990 - 31 March 1991 or 1 April 2001 - 30 November 2001, and zero otherwise.

There are six time-varying macro-economic covariates. Observations are taken quarterly and the values used in the estimation process are updated to the most recent quarterly value as each risk set is formed.

---

31 Most financial institutions are assigned investment grades for two reasons (Lando and Skodeberg, 2002). Firstly financial institutions are confidence and capital sensitive entities and it is difficult for them to operate with a poor credit profile or low credit rating. Speculative-grade rated financial institutions lack the strength necessary to compete and cling to life and such institutions tend to be weeded out as they merge or are acquired by other financial institutions to obtain more synergies. Secondly, banks have to observe regulatory constraints on the minimum risk capital and maintain sufficient required loss reserves. These factors contribute to making investment grade rating the norm for financial institutions.

Lando and Skodeberg (2002) separately considered rating observations of issuers in the financial sector. They found that the duration effect and downward momentum are less pronounced for financial institutions than for obligors in other sectors. For instance, rating states with the current rating grades varying from BB- to BBB- do not exhibit downward momentum. The focus of this thesis is on the impact of rating history on rating migrations, so it is sensible to exclude the financial institution sector from the study to avoid selectivity bias in the estimation process.

CFNAI (Chicago Fed National Activity Index): This is a composite index published by the Chicago Federal Reserve, which they compute as a three-month moving average of 85 monthly economic series.

RealGDPg actual minus potential (Actual real GDP growth minus potential real GDP growth or output growth gap): This variable measures deviation of the actual real GDP growth (as published by the US Bureau of Economic Analysis) from the potential real GDP growth (as published by the St. Louis Federal Reserve).

Industrial production change: as published by the US Federal Reserve Board.

S&P500 quarterly return: Definition sourced from Datastream.

S&P500 annualised SD (standard deviation): Daily returns for the quarter are used to compute the standard deviation and this is expressed as an annual standard deviation.

Term structure slope: The term structure slope is measured as the spread between US Treasury Constant Maturity three-month and 10-year rates as published by the US Federal Reserve.

4.2.2.3. Time interaction covariates

The TVC extended hazard model contains the 34 variables defined above plus 16 additional time interaction variables. In the estimation process the time interaction covariates $Z^i_p(t)$ between the rating history variable of state $i$ and event time $t$ are updated whenever an event occurs and a new risk set is formed. The 16 covariates that capture the interaction between the rating history variables of state $i$ and event time $t$ are constructed as follows:
\[
\begin{align*}
\text{Lag one time}_t^i &= \text{Lag one}^t * \text{Event time}_t^i, \\
\text{Lag two time}_t^i &= \text{Lag two}^t * \text{Event time}_t^i, \\
\text{Lag1 down time}_t^i &= \text{Dummy lag1 down}^t * \text{Event time}_t^i, \\
\text{Lag2 down time}_t^i &= \text{Dummy lag2 down}^t * \text{Event time}_t^i, \\
\text{Original rating time}_t^i &= \text{Original Rating}^t * \text{Event time}_t^i, \\
\text{Start rating time}_t^i &= \text{Start rating}^t * \text{Event time}_t^i, \\
\text{Investment boundary time}_t^i &= \text{Dummy investment boundary}^t * \text{Event time}_t^i, \\
\text{Junk boundary time}_t^i &= \text{Dummy junk boundary}^t * \text{Event time}_t^i, \\
\text{Not rated (NR) time}_t^i &= \text{Number NR}^t * \text{Event time}_t^i, \\
\text{Fallen Angel (FA) time}_t^i &= \text{Number FA}^t * \text{Event time}_t^i, \\
\text{Rising Star (RS) time}_t^i &= \text{Number RS}^t * \text{Event time}_t^i, \\
\text{Big down time}_t^i &= \text{Number big down}^t * \text{Event time}_t^i, \\
\text{Big up time}_t^i &= \text{Number big up}^t * \text{Event time}_t^i \\
\end{align*}
\]

The rating volatility measures, \textit{Rate of prior change} and \textit{Rate of prior down}, are declining functions of the calendar time since a firm entered the study. The effect of time passing is accounted for by updating the calendar age, thus:

\[
\begin{align*}
\text{Rate down time}_t^i &= \frac{\text{Number of prior downgrades}^t}{\text{Time spent in the study}^i_t}, \\
\text{Rate change time}_t^i &= \frac{\text{Number of prior rating changes}^t}{\text{Time spent in the study}^i_t} \\
\end{align*}
\]

\text{(17)}

Where:

\[
\text{Time spent in the study}^i_t = \text{Length of the time since the issuer entered the study till the start of rating state i} + \text{Event time}_t \\
\]

\text{(18)}

Rating age (\textit{Age since first rated}) is also a function of the calendar time since a firm was first rated. Any linear effects from increasing rating age as rating duration expands are automatically absorbed into the baseline hazard (Hosmer \textit{et al.}, 2008, p. 214). Thus, a nonlinear function of age and event time \(t\) is used as follows:

\[
\text{Log age time}^i_t = \text{Log (Age since first rated}^t + \text{Event time}) \\
\]

\text{(19)}
The time interaction covariates $Z_p^q(t)$ used to form forecasts for holdout state $q$ at forecast horizon $t$ are constructed similarly to the time interaction covariates for estimation state $i$ at event time $t$, $Z_p^i(t)$. However, the value of event time, in the equations (16)-(19) is now replaced by the conditional survival duration $T^*$ with $T^* < forecast horizon t$. In other words, for holdout state $q$, $Z_p^q(t) = Z_p^q(t = T^*)$. As $T^*$ can take different values over the forecast horizon $t$, the time interaction covariates $Z_p^q(t)$ will be updated as the conditional survival times $T^*$ is extended.

4.3. Data

4.3.1. Estimation and holdout periods

The year 1984 was chosen as the starting point of the study for two reasons. Rating migrations are more common events in the high-yield bond sector, and the high-yield bond market in the US was being substantially established in the first half of the 1980s. The growth of original issue junk bond and migrations from 1984, consequent to the establishment of the high-yield bond market, were expected to constitute a significant source of events for the study. Furthermore, macro-economic variables entered the models in the form of six quarters of distributed lags and 1982 is as far back as all data are available. Thus, the earliest possible starting point of the estimation period is mid-1983.

This thesis considers a 17-year estimation period to obtain a reasonably large starting pool for the estimation process. The time span of 17 years (1984-2000) is long enough to cover different phases of business cycles in the US, and captures major market downturns and international crises. The estimation period witnessed the US stock market crash in 1987, the Mexican currency crisis in 1994, the Thai financial crisis in

The period subsequent to the estimation period, 2001-2005, was used to construct a holdout sample. This period saw the Internet bubble burst and the 9/11 terrorist attack. The bursting of the Internet bubble was followed by an economic slowdown, falling business investment and negative stock returns for three consecutive years – 2000 to 2002. The year 2002 was considered the worst year for the corporate bond market in over 20 years. This year also saw unprecedented credit deterioration and the dramatic bankruptcies of fallen angels like WorldCom. Of the ten biggest bankruptcies in US history up to the end of the study (2005), seven, worth about $335 billion of assets, occurred in 2001-2002. Corporate rating volatility intensifies and the default rate escalated.

The time series for the exponentially weighted average of macro-economic variables used in this study are depicted in Figure 4-7 and Figure 4-8 below.
Figure 4-7

Performance of the US’s stock and money market
(Exponentially weighted average)

Figure 4-8

Performance of the US’s economy
(Exponentially weighted average except CFNAI)
Descriptive statistics of the time series for the exponentially weighted average of macro-economic variables are presented in Table 4-4 below.

Table 4-4

Descriptive statistics of macro-economic variables (exponentially weighted average except CFNAI)

<table>
<thead>
<tr>
<th>Macroeconomic Variable</th>
<th>Sample</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFNAI</td>
<td>1984-2000</td>
<td>0.113</td>
<td>0.447</td>
<td>-1.287</td>
<td>1.156</td>
<td>0.155</td>
<td>-0.677</td>
<td>1.711</td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>-0.301</td>
<td>0.466</td>
<td>-1.109</td>
<td>0.352</td>
<td>-0.353</td>
<td>-0.197</td>
<td>-1.087</td>
</tr>
<tr>
<td>RealGDPp actual minus potential</td>
<td>1984-2000</td>
<td>0.131</td>
<td>0.348</td>
<td>-0.927</td>
<td>1.126</td>
<td>0.153</td>
<td>0.045</td>
<td>2.623</td>
</tr>
<tr>
<td>(percentage)</td>
<td>2001-2005</td>
<td>-0.160</td>
<td>0.352</td>
<td>-0.795</td>
<td>0.320</td>
<td>-0.194</td>
<td>-0.258</td>
<td>-1.286</td>
</tr>
<tr>
<td>Industrial production change</td>
<td>1984-2000</td>
<td>0.939</td>
<td>0.662</td>
<td>-0.853</td>
<td>2.521</td>
<td>1.011</td>
<td>-0.108</td>
<td>0.071</td>
</tr>
<tr>
<td>(percentage)</td>
<td>2001-2005</td>
<td>0.183</td>
<td>0.655</td>
<td>-1.293</td>
<td>1.085</td>
<td>0.420</td>
<td>-1.020</td>
<td>0.138</td>
</tr>
<tr>
<td>SP500 quarterly return</td>
<td>1984-2000</td>
<td>3.609</td>
<td>2.990</td>
<td>-3.869</td>
<td>9.484</td>
<td>3.294</td>
<td>-0.077</td>
<td>-0.458</td>
</tr>
<tr>
<td>(percentage)</td>
<td>2001-2005</td>
<td>-0.501</td>
<td>4.333</td>
<td>-9.267</td>
<td>6.385</td>
<td>0.157</td>
<td>-0.344</td>
<td>-0.710</td>
</tr>
<tr>
<td>SP500 annual SD</td>
<td>1984-2000</td>
<td>1.769</td>
<td>0.612</td>
<td>0.649</td>
<td>3.789</td>
<td>1.678</td>
<td>1.014</td>
<td>1.303</td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>2.284</td>
<td>0.680</td>
<td>1.270</td>
<td>3.233</td>
<td>2.467</td>
<td>-0.279</td>
<td>-1.435</td>
</tr>
<tr>
<td>Term structure slope</td>
<td>1984-2000</td>
<td>1.716</td>
<td>0.943</td>
<td>0.165</td>
<td>3.363</td>
<td>1.718</td>
<td>-0.018</td>
<td>-1.299</td>
</tr>
<tr>
<td>(percentage)</td>
<td>2001-2005</td>
<td>2.140</td>
<td>1.029</td>
<td>-0.128</td>
<td>3.156</td>
<td>2.621</td>
<td>-1.200</td>
<td>0.248</td>
</tr>
</tbody>
</table>

As shown in Appendix A Table A-1, the difference in the mean/median macro-economic values between the estimation and holdout period is statistically significant at 1 percent level (except term structure slope).

The holdout period 2001-2005 was markedly different from the estimation period 1984-2000. It is therefore not surprising that the holdout period, compared with the estimation period, experienced a different rating migration pattern, had a markedly different rating history and current rating profile.

4.3.2. Estimation and holdout samples

Rating data was obtained from Standard & Poor’s CreditPro2005. The data cover 25 years of rating history of issuers rated by Standard & Poor’s at any time between 1 January 1981 and 31 December 2005. The dataset includes the issuer rating history of 11,605 corporate obligors. There is an overwhelming bias towards issuers domiciled in the US - 63.2 percent of total observations relate to US issuers. Inclusion of foreign
issuers in the study was considered, but sample sizes across countries were often small and were highly variable over time. Early experimentation with the data suggested that there was a country effect, but it was unclear how much this was driven by real effects and how much by idiosyncratic variation in sample size by country. In the interest of greater homogeneity in the sample and keeping the number of variables in the model manageable, the sample was restricted to US data.

The estimation sample covering the period 1984-2000 was formed from 4487 available rating states of US non-financial issuers, excluding observations which did not pass the screening test of experiencing at least two migrations. The holdout sample (2001-2005) included 1872 available rating states of US non-financial obligors. Table 4-5 describes the process used to construct the estimation and holdout samples.

| Table 4-5 |
|---|---|
| **The construction of estimation and holdout samples** |  |
| Pooled observations across issuers and rating grades, within the period analysed | 19315 | 15308 |
| After excluding rating states being censored or unrated (NR) at lag two rating states | 6234 | 3373 |
| After excluding rating states of firms from countries other than the US | 5187 | 1934 |
| After excluding rating states of US firms in the Sector Financial Institutions | **4487** | **1872** |

The distribution of rating states in the estimation sample categorised by the *start rating* and the beginning year are depicted in Table 4-6. There is a spread of observations across the rating spectrum and across time. However, there is a bias towards issuers in the low end investment-grade and upper speculative-grade ratings. Furthermore, rating states commenced in the 1990s dominate the estimation sample.

### 4.3.3. Rating migrations

The distributions of rating states in the estimation sample categories by *start rating* and *end rating* are depicted in Table 4-7.
### Table 4-6

**DISTRIBUTION OF RATING STATES BY START RATING AND START YEAR, 1984-2000**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (C )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0.16%</td>
<td></td>
</tr>
<tr>
<td>5 (CC)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>22</td>
<td>22</td>
<td>68</td>
<td>1.52%</td>
<td></td>
</tr>
<tr>
<td>7 (CCC-)</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>15</td>
<td>16</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>11</td>
<td>12</td>
<td>113</td>
<td>2.52%</td>
<td></td>
</tr>
<tr>
<td>8 (CCC)</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>24</td>
<td>27</td>
<td>128</td>
<td>2.87%</td>
<td></td>
</tr>
<tr>
<td>9 (CCC+)</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>13</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>18</td>
<td>20</td>
<td>144</td>
<td>3.21%</td>
<td></td>
</tr>
<tr>
<td>10 (B-)</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>18</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>11</td>
<td>24</td>
<td>30</td>
<td>172</td>
<td>3.83%</td>
<td></td>
</tr>
<tr>
<td>11 (B)</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>17</td>
<td>2</td>
<td>11</td>
<td>17</td>
<td>11</td>
<td>8</td>
<td>15</td>
<td>6</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>22</td>
<td>33</td>
<td>193</td>
<td>4.30%</td>
<td></td>
</tr>
<tr>
<td>12 (B+)</td>
<td>2</td>
<td>12</td>
<td>14</td>
<td>18</td>
<td>15</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>17</td>
<td>26</td>
<td>9</td>
<td>13</td>
<td>19</td>
<td>17</td>
<td>25</td>
<td>24</td>
<td>26</td>
<td>287</td>
<td>6.40%</td>
<td></td>
</tr>
<tr>
<td>13 (BB- )</td>
<td>1</td>
<td>7</td>
<td>17</td>
<td>10</td>
<td>12</td>
<td>26</td>
<td>12</td>
<td>18</td>
<td>23</td>
<td>24</td>
<td>12</td>
<td>13</td>
<td>24</td>
<td>17</td>
<td>23</td>
<td>18</td>
<td>26</td>
<td>283</td>
<td>6.31%</td>
<td></td>
</tr>
<tr>
<td>14 (BB)</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>18</td>
<td>17</td>
<td>28</td>
<td>15</td>
<td>19</td>
<td>22</td>
<td>34</td>
<td>12</td>
<td>20</td>
<td>19</td>
<td>13</td>
<td>22</td>
<td>27</td>
<td>31</td>
<td>340</td>
<td>6.78%</td>
<td></td>
</tr>
<tr>
<td>15 (BB+)</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>19</td>
<td>15</td>
<td>19</td>
<td>11</td>
<td>14</td>
<td>18</td>
<td>28</td>
<td>15</td>
<td>23</td>
<td>18</td>
<td>25</td>
<td>29</td>
<td>23</td>
<td>24</td>
<td>286</td>
<td>6.37%</td>
<td></td>
</tr>
<tr>
<td>16 (BBB-)</td>
<td>0</td>
<td>2</td>
<td>24</td>
<td>21</td>
<td>18</td>
<td>25</td>
<td>32</td>
<td>19</td>
<td>30</td>
<td>14</td>
<td>28</td>
<td>17</td>
<td>35</td>
<td>32</td>
<td>37</td>
<td>25</td>
<td>37</td>
<td>396</td>
<td>8.83%</td>
<td></td>
</tr>
<tr>
<td>17 (BBB)</td>
<td>2</td>
<td>14</td>
<td>25</td>
<td>19</td>
<td>25</td>
<td>28</td>
<td>26</td>
<td>29</td>
<td>19</td>
<td>30</td>
<td>33</td>
<td>26</td>
<td>42</td>
<td>40</td>
<td>22</td>
<td>40</td>
<td>42</td>
<td>462</td>
<td>10.30%</td>
<td></td>
</tr>
<tr>
<td>18 (BBB+)</td>
<td>3</td>
<td>9</td>
<td>23</td>
<td>14</td>
<td>24</td>
<td>28</td>
<td>20</td>
<td>22</td>
<td>19</td>
<td>23</td>
<td>26</td>
<td>29</td>
<td>28</td>
<td>19</td>
<td>30</td>
<td>44</td>
<td>33</td>
<td>394</td>
<td>8.78%</td>
<td></td>
</tr>
<tr>
<td>19 (A-)</td>
<td>0</td>
<td>9</td>
<td>17</td>
<td>20</td>
<td>23</td>
<td>15</td>
<td>27</td>
<td>22</td>
<td>24</td>
<td>23</td>
<td>20</td>
<td>21</td>
<td>33</td>
<td>24</td>
<td>20</td>
<td>24</td>
<td>25</td>
<td>347</td>
<td>7.73%</td>
<td></td>
</tr>
<tr>
<td>20 (A)</td>
<td>0</td>
<td>14</td>
<td>19</td>
<td>4</td>
<td>13</td>
<td>21</td>
<td>25</td>
<td>29</td>
<td>32</td>
<td>23</td>
<td>18</td>
<td>13</td>
<td>26</td>
<td>18</td>
<td>23</td>
<td>23</td>
<td>18</td>
<td>25</td>
<td>332</td>
<td>7.40%</td>
</tr>
<tr>
<td>21 (A+)</td>
<td>2</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>13</td>
<td>17</td>
<td>15</td>
<td>12</td>
<td>18</td>
<td>11</td>
<td>20</td>
<td>29</td>
<td>11</td>
<td>21</td>
<td>21</td>
<td>13</td>
<td>32</td>
<td>258</td>
<td>5.75%</td>
<td></td>
</tr>
<tr>
<td>22 (AA-)</td>
<td>0</td>
<td>5</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>14</td>
<td>7</td>
<td>14</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>22</td>
<td>155</td>
<td>3.45%</td>
<td></td>
</tr>
<tr>
<td>23 (AA)</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>104</td>
<td>2.32%</td>
<td></td>
</tr>
<tr>
<td>24 (AA+)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>28</td>
<td>0.62%</td>
</tr>
<tr>
<td>26 (AAA)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>0.56%</td>
<td></td>
</tr>
</tbody>
</table>

| Total | 19 | 100 | 222 | 197 | 234 | 275 | 285 | 275 | 292 | 275 | 245 | 257 | 318 | 303 | 323 | 402 | 465 | 4487 | 100% |
| Percentile | 0.42% | 2.23% | 4.95% | 4.39% | 5.22% | 6.13% | 6.35% | 6.13% | 6.51% | 6.13% | 5.46% | 5.73% | 7.09% | 6.75% | 7.20% | 8.96% | 10.36% | 100% |
Table 4-7

**DISTRIBUTION OF RATING STATES BY START RATING AND END RATING, 1984-2000**

<table>
<thead>
<tr>
<th>End rating</th>
<th>Not rated (NR)</th>
<th>Default</th>
<th>C</th>
<th>CC</th>
<th>CCC-</th>
<th>CCC</th>
<th>CCC+</th>
<th>B-</th>
<th>B</th>
<th>B+</th>
<th>BB-</th>
<th>BB</th>
<th>BB+</th>
<th>BBB-</th>
<th>BBB</th>
<th>BBB+</th>
<th>A-</th>
<th>A</th>
<th>A+</th>
<th>AA-</th>
<th>AA</th>
<th>AA+</th>
<th>AAA</th>
<th>Total</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start rating</td>
<td>2 (C )</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5 (CC)</td>
<td>1</td>
<td>59</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>68</td>
</tr>
<tr>
<td>7 (CCC-)</td>
<td>11</td>
<td>79</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>113</td>
</tr>
<tr>
<td>8 (CCC)</td>
<td>15</td>
<td>55</td>
<td>1</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>129</td>
<td>2.87%</td>
</tr>
<tr>
<td>9 (CCC+)</td>
<td>10</td>
<td>51</td>
<td>1</td>
<td>14</td>
<td>18</td>
<td>14</td>
<td>16</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>10 (B-)</td>
<td>21</td>
<td>22</td>
<td>0</td>
<td>10</td>
<td>13</td>
<td>31</td>
<td>15</td>
<td>29</td>
<td>12</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>172</td>
<td>3.83%</td>
<td></td>
</tr>
<tr>
<td>11 (B)</td>
<td>32</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>27</td>
<td>20</td>
<td>33</td>
<td>31</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>193</td>
<td>4.30%</td>
</tr>
<tr>
<td>12 (B+)</td>
<td>50</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>27</td>
<td>47</td>
<td>46</td>
<td>55</td>
<td>23</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>287</td>
<td>6.40%</td>
</tr>
<tr>
<td>13 (BB-)</td>
<td>55</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>10</td>
<td>51</td>
<td>47</td>
<td>70</td>
<td>21</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>283</td>
<td>6.31%</td>
</tr>
<tr>
<td>14 (BB)</td>
<td>40</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>21</td>
<td>54</td>
<td>53</td>
<td>78</td>
<td>36</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>304</td>
<td>6.78%</td>
</tr>
<tr>
<td>15 (BB+)</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>24</td>
<td>39</td>
<td>52</td>
<td>101</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>286</td>
<td>6.37%</td>
<td></td>
</tr>
<tr>
<td>16 (BBB-)</td>
<td>60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>38</td>
<td>70</td>
<td>94</td>
<td>94</td>
<td>18</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>396</td>
<td>8.83%</td>
</tr>
<tr>
<td>17 (BBB)</td>
<td>85</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>14</td>
<td>103</td>
<td>107</td>
<td>100</td>
<td>28</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>462</td>
<td>10.30%</td>
<td></td>
</tr>
<tr>
<td>18 (BBB+)</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>15</td>
<td>109</td>
<td>102</td>
<td>79</td>
<td>16</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>394</td>
<td>8.78%</td>
</tr>
<tr>
<td>19 (A-)</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>41</td>
<td>82</td>
<td>77</td>
<td>87</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>347</td>
<td>7.73%</td>
<td></td>
</tr>
<tr>
<td>20 (A)</td>
<td>47</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>33</td>
<td>65</td>
<td>94</td>
<td>71</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>332</td>
<td>7.40%</td>
<td></td>
</tr>
<tr>
<td>21 (A+)</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>67</td>
<td>94</td>
<td>30</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>258</td>
<td>5.75%</td>
</tr>
<tr>
<td>22 (AA-)</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>54</td>
<td>47</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>155</td>
</tr>
<tr>
<td>23 (AA)</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>33</td>
<td>27</td>
<td>9</td>
<td>3</td>
<td>104</td>
<td>2.32%</td>
</tr>
<tr>
<td>24 (AA+)</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>14</td>
<td>2</td>
<td>8</td>
<td>0.62%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 (AAA)</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>25</td>
<td>0.56%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total: 646 299 6 48 66 85 77 106 126 183 207 251 249 360 399 351 265 287 245 121 61 33 16 4487 100%

Percentile: 14.40% 6.66% 0.13% 1.07% 1.47% 1.89% 1.72% 2.36% 2.81% 4.08% 4.61% 5.59% 5.55% 8.02% 8.89% 7.82% 5.91% 6.40% 5.46% 2.70% 1.36% 0.74% 0.36% 100%
Figure 4-9

The distribution of rating changes and rating withdrawals by year of event, 1984-2000

Figure 4-10

The distribution of rating changes and rating withdrawals by year of event, 2001-2005
Figure 4-11

The distribution of rating changes and rating withdrawals by start rating 1984-2000

![Chart showing distribution of rating changes and rating withdrawals by start rating from 1984 to 2000.]

Figure 4-12

The distribution of rating changes and rating withdrawals by start rating 2001-2005

![Chart showing distribution of rating changes and rating withdrawals by start rating from 2001 to 2005.]

Legend:
- Blue: The number of upgrades per rating class
- Red: The number of downgrades per rating class
- Green: The number of unrated (NR) states per rating class
A striking feature is that multiple notch-rating changes are infrequently observed. Figures 4-9 and 4-10, which follow Table 4-7, depict the distribution of rating changes and rating withdrawals during the period 1984-2000 and 2001-2005. There were substantially more downgrades than upgrades in both estimation and holdout samples. Ratings were less stable and were more frequently downgraded than upgraded in recent years. Downgrades increased sharply in the late 1990s and early 2000s.33

In the estimation period (Figure 4-9) the downgrade incidence is 0.45 times higher than the upgrade incidence, whereas in the holdout period (Figure 4-10), this increases to 2.27 times. The number of downgrade (upgrade) occurrences in the five-year holdout period (Figure 4-10) is equivalent to 47.17 percent (20.85 percent) of the respective migration incidences in the 17-year estimation period (Figure 4-9). Compared with the migration pattern in the estimation period, the proportion of upgrades has dropped sharply and the proportion of downgrades has risen substantially in the holdout period. As shown in Appendix A Table A-2, the proportion of upgrades/ downgrades in the holdout period is statistically different from the respective proportion in the estimation period at 1 percent significant level.

Histograms of rating changes and rating withdrawals (NR) categorised by the start rating are depicted in Figure 4-11 (estimation period) and Figure 4-12 (holdout period) above. The rating migration dynamics vary between the estimation and the holdout samples. Upgrades in the estimation period are heavily concentrated at the start rating spectrum of B+ to A, whereas upgrades in the holdout period mass at the

33 The substantially fewer migrations and rating withdrawals in 2001 reflect the truncation of rating data in forming the holdout sample on 1 January 2001.
speculative grade ratings of CCC to BB-. There is a spread of downgrades across a wide rating spectrum of CC to AA- in the estimation period while downgrades in the holdout period are concentrated on the speculative grade ratings of CC to BB.

As shown in Figure 4-11, 37.5 percent of the downgrades and 60.25 percent of the upgrades observed in the period 1984-2000 are from the investment/speculative rating thresholds. There is a shift towards lower start ratings for both down states and up states in the holdout sample. Figure 4-12 shows that 30.71 percent of the downgrades and 31.43 percent of the upgrades observed in the period 2001-2005 are from the boundary ratings.

Figures 4-13 and 4-14 below depict the histogram of downgrade and upgrade changes during the estimation period, 1984-2000. As can be seen from Figure 4-13 and 4-14, the majority of rating transitions are to the neighbouring state (within a single rating notch).
Figure 4-13

Histogram of downgrade changes, 1984-2000

Figure 4-14

Histogram of upgrade changes, 1984-2000
Figure 4-15

Histogram of time to events

Time to Down—grades Histogram, 1984—2000

Time to Up—grades Histogram, 1984—2000

Time to Down—grades Histogram, 2001—2005

Time to Up—grades Histogram, 2001—2005
4.3.4. Time to events (Survival durations)

Histograms of the time to downgrades and the time to upgrades (survival duration) for estimation and holdout states are depicted in Figure 4-15 above. The scale of the time axes differ between the estimation and holdout samples since most of the downgrade states do not last a year in the holdout period.

Both upgrades and downgrades have positively skewed distributions. The range of the distributions is similar, but it is clear that durations for downgrades tend to be shorter than for upgrades. There is a noticeable concentration of downgrades in durations shorter than one year. A striking feature in Figure 4-15 is that upgrades in the holdout period mass over a wider range of duration – from a quarter to two and a half years.

Table 4-8 provides descriptive statistics for the survival durations of rating states in the estimation and holdout samples.

Table 4-8

<table>
<thead>
<tr>
<th>Samples</th>
<th>Rating states</th>
<th>Number of observations</th>
<th>Mean (years)</th>
<th>Standard Deviation</th>
<th>Min (days)</th>
<th>Max (years)</th>
<th>Median (years)</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984-2000</td>
<td>Down states</td>
<td>1698</td>
<td>1.658</td>
<td>1.957</td>
<td>1</td>
<td>15.345</td>
<td>0.997</td>
<td>2.365</td>
<td>7.547</td>
</tr>
<tr>
<td></td>
<td>Up states</td>
<td>1175</td>
<td>2.247</td>
<td>1.837</td>
<td>6</td>
<td>11.436</td>
<td>1.715</td>
<td>1.837</td>
<td>4.220</td>
</tr>
<tr>
<td></td>
<td>All states</td>
<td>4487</td>
<td>2.129</td>
<td>2.270</td>
<td>1</td>
<td>15.967</td>
<td>1.416</td>
<td>2.115</td>
<td>5.688</td>
</tr>
<tr>
<td>2001-2005</td>
<td>Down states</td>
<td>801</td>
<td>0.514</td>
<td>0.596</td>
<td>1</td>
<td>4.337</td>
<td>0.296</td>
<td>2.378</td>
<td>8.074</td>
</tr>
<tr>
<td></td>
<td>Up states</td>
<td>245</td>
<td>1.202</td>
<td>0.840</td>
<td>1</td>
<td>4.690</td>
<td>1.189</td>
<td>0.749</td>
<td>0.599</td>
</tr>
<tr>
<td></td>
<td>All states</td>
<td>1872</td>
<td>0.944</td>
<td>0.9289</td>
<td>1</td>
<td>4.604</td>
<td>0.6425</td>
<td>1.2936</td>
<td>1.1187</td>
</tr>
</tbody>
</table>

Of 4487 estimation rating states, 1698 states (37.84 percent) experienced downgrades, and 1175 states (26.19 percent) experienced upgrades. Downgrade durations vary from 1 day to 15.35 years, have a mean length of 1.66 years, and a median length of 0.99 years. Upgrade durations vary from 6 days to 11.44 years, have a mean length of 2.25 years, and a median length of 1.72 years.
Of 1872 holdout states, downgrades represent 42.79 percent and upgrades represent 13.09 percent of the samples. As presented in Appendix A Table A-3, the descriptive statistics for the survival durations of all states, down states, and up states in the estimation and holdout periods are statistically different at 1 percent significance level.

**4.3.5. Start rating and rating history variables**

**4.3.5.1. Start rating**

The histogram of *start rating* for observations in the estimation and holdout samples are depicted in Figures 4-16 and 4-17. Start ratings for estimation states are massed at the middle rating grades varying from 12 (B+) to 20 (A), whereas start ratings for holdout states are concentrated in the region of speculative rating grades, varying from 8 (CCC) to 14 (BB).

Descriptive statistics of the current rating (*start rating*) for down states, up states, and all states in the estimation and holdout samples are given in Table 4-9.

**Table 4-9**

<table>
<thead>
<tr>
<th>Samples</th>
<th>Rating states</th>
<th>Number of observations</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984-2000 Down-states</td>
<td>1698</td>
<td>14.715</td>
<td>4.958</td>
<td>2 ( C )</td>
<td>26 (AAA)</td>
<td>16 (BBB-)</td>
<td>-0.2140</td>
<td>-0.9699</td>
<td></td>
</tr>
<tr>
<td>Up-states</td>
<td>1175</td>
<td>15.570</td>
<td>3.398</td>
<td>2 ( C )</td>
<td>24 (AA+)</td>
<td>16 (BBB-)</td>
<td>-0.2649</td>
<td>-0.0274</td>
<td></td>
</tr>
<tr>
<td>All states</td>
<td>4487</td>
<td>15.610</td>
<td>4.351</td>
<td>2 ( C )</td>
<td>26 (AAA)</td>
<td>16 (BBB-)</td>
<td>-0.333</td>
<td>-0.416</td>
<td></td>
</tr>
<tr>
<td>2001-2005 Down-states</td>
<td>801</td>
<td>11.599</td>
<td>4.174</td>
<td>2 ( C )</td>
<td>24 (AA+)</td>
<td>11 (B)</td>
<td>0.3731</td>
<td>-0.4396</td>
<td></td>
</tr>
<tr>
<td>Up-states</td>
<td>245</td>
<td>11.706</td>
<td>3.660</td>
<td>5 (CC)</td>
<td>21 (A+)</td>
<td>11 (B)</td>
<td>0.3051</td>
<td>-0.1110</td>
<td></td>
</tr>
<tr>
<td>All states</td>
<td>1872</td>
<td>12.723</td>
<td>4.151</td>
<td>2 ( C )</td>
<td>26 (AAA)</td>
<td>12 (B+)</td>
<td>0.216</td>
<td>-0.457</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Appendix A Table A-4, the difference in the mean *start ratings* / the median *start ratings* between estimation and holdout states is statistically different at 1 percent significant level.
Figure 4-16

Histogram of start rating, 1984—2000

Figure 4-17

Histogram of start rating, 2001—2005
4.3.5.2. Rating history variables

The descriptive statistics of rating history for states in the estimation and holdout samples are given in Table 4-10.

### Table 4-10

Descriptive statistics of rating history

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
<th>Median</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age since first rated (year)</td>
<td>1984-2000</td>
<td>8.578</td>
<td>4.892</td>
<td>0.134</td>
<td>20.008</td>
<td>7.7342</td>
<td>-0.600</td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>10.257</td>
<td>7.331</td>
<td>0.203</td>
<td>24.992</td>
<td>7.3932</td>
<td>0.724</td>
</tr>
<tr>
<td>Original rating</td>
<td>1984-2000</td>
<td>17.254</td>
<td>4.385</td>
<td>7 (CCC-)</td>
<td>26 (AAA)</td>
<td>17 (BBB)</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>16.026</td>
<td>4.236</td>
<td>7 (CCC-)</td>
<td>26 (AAA)</td>
<td>15 (BBB)</td>
<td>0.507</td>
</tr>
<tr>
<td>Lag one (year)</td>
<td>1984-2000</td>
<td>2.131</td>
<td>2.059</td>
<td>0.003</td>
<td>16.126</td>
<td>1.5343</td>
<td>2.206</td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.966</td>
<td>0.917</td>
<td>0.003</td>
<td>4.690</td>
<td>0.6822</td>
<td>1.396</td>
</tr>
<tr>
<td>Dummy lag1 down</td>
<td>1984-2000</td>
<td>0.660</td>
<td>0.490</td>
<td>0</td>
<td>1</td>
<td>0.548</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.773</td>
<td>0.419</td>
<td>0</td>
<td>1</td>
<td>-1.304</td>
<td></td>
</tr>
<tr>
<td>Lag two (year)</td>
<td>1984-2000</td>
<td>2.183</td>
<td>2.025</td>
<td>0.003</td>
<td>16.833</td>
<td>1.6274</td>
<td>2.387</td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.981</td>
<td>0.830</td>
<td>0.003</td>
<td>4.718</td>
<td>0.7795</td>
<td>1.252</td>
</tr>
<tr>
<td>Dummy lag2 down</td>
<td>1984-2000</td>
<td>0.660</td>
<td>0.490</td>
<td>0</td>
<td>1</td>
<td>-0.406</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.823</td>
<td>0.382</td>
<td>0</td>
<td>1</td>
<td>-1.691</td>
<td></td>
</tr>
<tr>
<td>Rate prior change</td>
<td>1984-2000</td>
<td>0.640</td>
<td>0.536</td>
<td>0.061</td>
<td>14.898</td>
<td>0.5143</td>
<td>7.316</td>
</tr>
<tr>
<td>(number of migrations per year)</td>
<td>2001-2005</td>
<td>1.473</td>
<td>2.351</td>
<td>0.216</td>
<td>48.667</td>
<td>1.0965</td>
<td>15.048</td>
</tr>
<tr>
<td>Rate prior down</td>
<td>1984-2000</td>
<td>0.413</td>
<td>0.539</td>
<td>0</td>
<td>14.898</td>
<td>0.2841</td>
<td>7.601</td>
</tr>
<tr>
<td>(number of downgrades per year)</td>
<td>2001-2005</td>
<td>1.250</td>
<td>2.007</td>
<td>0</td>
<td>48.667</td>
<td>0.9395</td>
<td>15.758</td>
</tr>
<tr>
<td>Number NR</td>
<td>1984-2000</td>
<td>0.077</td>
<td>0.283</td>
<td>0</td>
<td>2</td>
<td>3.785</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.017</td>
<td>0.134</td>
<td>0</td>
<td>2</td>
<td>8.118</td>
<td></td>
</tr>
<tr>
<td>Number FA</td>
<td>1984-2000</td>
<td>0.262</td>
<td>0.485</td>
<td>0</td>
<td>2</td>
<td>1.627</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.274</td>
<td>0.456</td>
<td>0</td>
<td>2</td>
<td>1.149</td>
<td></td>
</tr>
<tr>
<td>Number RS</td>
<td>1984-2000</td>
<td>0.199</td>
<td>0.431</td>
<td>0</td>
<td>2</td>
<td>1.989</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.058</td>
<td>0.238</td>
<td>0</td>
<td>2</td>
<td>4.034</td>
<td></td>
</tr>
<tr>
<td>Number big down</td>
<td>1984-2000</td>
<td>0.301</td>
<td>0.565</td>
<td>0</td>
<td>4</td>
<td>2.037</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.342</td>
<td>0.598</td>
<td>0</td>
<td>4</td>
<td>1.735</td>
<td></td>
</tr>
<tr>
<td>Number big up</td>
<td>1984-2000</td>
<td>0.394</td>
<td>0.668</td>
<td>0</td>
<td>4</td>
<td>1.824</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2001-2005</td>
<td>0.339</td>
<td>0.388</td>
<td>0</td>
<td>3</td>
<td>2.960</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Appendix A Table A-5, the *t tests* indicate that, except for the variable *Number FA*, the means of rating history variables for estimation and holdout states are statistically different at 1 percent significance level. The Wilcoxon test also suggests that the median of rating history variables for estimation and holdout states are statistically different at 5 percent level significance.

Compared to the estimation period, the holdout period had a slightly higher percentage of fallen angel (FA) events and a substantially lower percentage of rising...
star (RS) events. In addition, downward momentum was observed more often in the holdout period than in the estimation period. Appendix A Table A-6 show that the proportions of estimation and holdout states experiencing a rising star, and downward momentum at both lag one and lag two are statistically different at 1 percent significance level. Similar analysis confirmed that the proportions of estimation and holdout states undergoing a fallen angel are statistically different at 5 percent significance level.

Figure 4-18 below depicts the histogram of rating states in the estimation and holdout samples categorised by the original rating. Figures 4-19 and 4-20 show the distribution of estimation and holdout rating states categorised by the year first rated. The study is dominated by seasoned issuers, with the median rating age of over seven years. As shown in Figure 4-19 and 4-20, the majority of issuers in the estimation sample were first rated in 1980. Whereas a substantial proportion of issuers in the holdout sample were first rated in 1980, or within the period 1997-2001.
Figure 4-18

Distribution of rating states by the original rating

Figure 4-19

Distribution of rating states in the estimation sample by the year first rated

Figure 4-20

Distribution of rating states in the holdout sample by the year first rated
4.4. Conclusion

Given the attractive features and the relevance to the objectives of the thesis, Cox’s hazard models (Cox, 1972) have been used to investigate the effects of rating history and time on rating migration hazards in both static and dynamic estimation frameworks. The approach proposed by Andersen (1992) and Chen et al. (2005) has been applied to estimate integrated baseline hazard function and generate dynamic survival probability forecasts in the presence of time-varying covariates.

Three stratified Cox’s hazard models were developed for each down state and up state case in the estimation sample. Downgrades and upgrades were modelled as competitive risk. In forming the downgrade model, upgrade observations were treated as censored and vice versa. The conditional gap time approach (Prentice et al., 1981) utilised in the estimation approach accounts for the time sequence of recurrent migration events each rating state experienced since the beginning of the observation period.

The stratified Cox’s proportional hazard model (time-fixed model) employs 16 time-fixed rating history variables, 11 industry sector dummies, and 7 time-fixed macro-economic covariates. The time-fixed model only accounts for the macro-economic conditions measured at the beginning of each rating state. The stratified Cox’s time-varying covariate hazard base model (TVC base model) differs from the time-fixed model in that it incorporates one time-fixed macro-economic indicator, and six time-varying macro-economic covariates, updated prior to each event time. The TVC base model therefore captures the macro-economic stage at the beginning of each rating state and the evolution of macro-economic conditions over rating durations. The stratified Cox’s TVC hazard extended model (TVC extended model) extends the TVC base model. Sixteen time-varying covariates that capture the interactions between
rating history variables and time were added. Relative to the TVC base model, the
TVC extended model accounts for not only the evolution of macro-economic
conditions but also the interaction between rating history and the passage of time over
rating duration.

Given the positive features over the measures commonly used to assess the
discrimination power of a rating system, the Brier score (Brier, 1950) and its
covariance decompositions (Yates, 1982) have been used to assess various attributes
of probability survival estimates generated by the time-fixed model, and the TVC base and TVC extended models at different time horizons.

Rating history data from Standard & Poor’s CreditPro2005 dataset were used in
subsequent chapters to examine non-Markovian behaviours in the rating dynamics of
US non-financial firms during the period 1984-2000. The forecast accuracy
assessment of the time-fixed and TVC hazard models has been examined on a holdout sample that includes rating states of US non-financial firms in the period 2001-2005.

The macro-economic conditions in the holdout period were much worse than the
conditions in the estimation period. Consequently, relative to the estimation period,
the holdout period observed a markedly different rating migration pattern and rating history profile. It is characterised by higher rating volatility and a lower current rating profile. The holdout period experienced more downgrade and fallen angel occurrences and fewer upgrade and rising star incidences. In addition, it took much less for a downgrade to occur in the holdout period than in the estimation period.
Chapter 5

Rating migrations: The effects of rating history

This chapter develops a stratified Cox’s proportional hazard model, as in equation (2), to estimate the hazard of an upgrade and a downgrade. The objective is twofold. First, to investigate non-Markovian behaviours, specifically the impact of rating history on rating migration dynamics. This is discussed in section 5.1 below. Second, to investigate the predictive accuracy of rating history when forecasting future rating migrations, as discussed in section 5.2.

5.1. Estimation results

The stratified Cox’s proportional hazard models (time-fixed models) incorporate covariates whose values were measured at the beginning of each rating state. These covariates enter equation (2) without being updated as time in a rating state extends. There are a large number of potential covariates to include in the model, 34 in all. In the interest of a parsimonious model and compact presentation of the results, a backward stepwise estimation procedure was employed. Variables were retained in the models according to the log-likelihood ratio test if they were significant at the 10 percent level or better.

The significant coefficients of the time-fixed models are given for downgrades and upgrades in Panel A of Table 5-1. Panel B provides statistics on the significance of the overall models. These statistics show that the model is strongly significant. Panel C and Panel D summarise the number of rating migrations and censored states observed across the strata of multiple events during the estimation and holdout
periods respectively. The maximum number of rating events which a firm experienced during the estimation and holdout periods is 14 (Panel C) and 7 (Panel D) respectively. Panel C shows that most estimation firms underwent from 1 to 6 migrations and there are few observations in the strata of 8 to 14 repeated events. Panel D indicates that the majority of holdout firms experienced from 1 to 4 migrations and few are observed in the strata with 5 to 7 repeated events.

Table 5-1

STRATIFIED COX'S PROPORTIONAL HAZARD (TIME-FIXED) MODELS
PANEL A: COEFFICIENT ESTIMATES (BACKWARD SELECTION), 1984-2000

<table>
<thead>
<tr>
<th>Variables</th>
<th>Downgrade model</th>
<th>Upgrade model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Rating history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age_since_first_rated</td>
<td>0.03691***</td>
<td>0.01092</td>
</tr>
<tr>
<td>Original_rating</td>
<td>-0.10706***</td>
<td>0.00814</td>
</tr>
<tr>
<td>Start_rating</td>
<td>-0.33778***</td>
<td>0.07102</td>
</tr>
<tr>
<td>Dummy_inv_boundary</td>
<td>-0.32807***</td>
<td>0.08319</td>
</tr>
<tr>
<td>Lag_one</td>
<td>-0.08068***</td>
<td>0.01906</td>
</tr>
<tr>
<td>Dummy_lag1_down</td>
<td>0.94341***</td>
<td>0.06321</td>
</tr>
<tr>
<td>Lag_two</td>
<td>-0.06177***</td>
<td>0.01847</td>
</tr>
<tr>
<td>Dummy_lag2_down</td>
<td>0.20739***</td>
<td>0.06285</td>
</tr>
<tr>
<td>Rate_prior_change</td>
<td>0.14321**</td>
<td>0.06762</td>
</tr>
<tr>
<td>Number_NR</td>
<td>-0.4924***</td>
<td>0.11963</td>
</tr>
<tr>
<td>Number_FA</td>
<td>-0.41309***</td>
<td>0.08678</td>
</tr>
<tr>
<td>Number_RS</td>
<td>0.27749**</td>
<td>0.10776</td>
</tr>
<tr>
<td>Number_big_down</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number_big_up</td>
<td>0.092*</td>
<td>0.0538</td>
</tr>
<tr>
<td>Macro-economic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy_recession</td>
<td></td>
<td>-0.34342**</td>
</tr>
<tr>
<td>CFNAI</td>
<td>-0.24371*</td>
<td>0.12814</td>
</tr>
<tr>
<td>RealGDPg_actual_minus_potential</td>
<td>0.56883***</td>
<td>0.18401</td>
</tr>
<tr>
<td>Industrial_production_change</td>
<td>0.04174***</td>
<td>0.01097</td>
</tr>
<tr>
<td>SP500_quarterly_return</td>
<td>-0.42732***</td>
<td>0.14616</td>
</tr>
<tr>
<td>SP500_annual_SD</td>
<td>0.2276***</td>
<td>0.04501</td>
</tr>
<tr>
<td>Term_structure_slope</td>
<td>-0.1822***</td>
<td>0.031</td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerospace/automotive/capital goods/metals</td>
<td>-0.17056**</td>
<td>0.0774</td>
</tr>
<tr>
<td>Consumer / service sector</td>
<td>-0.67424***</td>
<td>0.12622</td>
</tr>
<tr>
<td>Energy and natural resources</td>
<td>-0.46793***</td>
<td>0.12374</td>
</tr>
<tr>
<td>Telecommunications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure time / media</td>
<td>-0.43272***</td>
<td>0.14616</td>
</tr>
<tr>
<td>Real Estate</td>
<td>-2.57232***</td>
<td>0.87784</td>
</tr>
<tr>
<td>Utility</td>
<td>-0.25932***</td>
<td>0.07938</td>
</tr>
<tr>
<td>Forest and building products/homebuilders</td>
<td>-0.80988***</td>
<td>0.15085</td>
</tr>
<tr>
<td>Health care / chemicals</td>
<td>-0.18775</td>
<td>0.12105</td>
</tr>
<tr>
<td>Transportation</td>
<td>-0.27493**</td>
<td>0.12821</td>
</tr>
<tr>
<td>High technology/computers/office equipment</td>
<td>-0.25115*</td>
<td>0.14603</td>
</tr>
</tbody>
</table>

*** P-value ≤ 1%, ** 1%< P-value ≤ 5%, * 5%< P-value ≤ 10% based on Wald chi-square tests
### Table 5-1 (continued)

#### STRATIFIED COX’S PROPORTIONAL HAZARD (TIME-FIXED) MODELS

**PANEL B: TESTING GLOBAL NULL HYPOTHESIS $\beta=0$, 1984-2000**

<table>
<thead>
<tr>
<th></th>
<th>Downgrade model</th>
<th></th>
<th>Upgrade model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Likelihood Ratio</td>
<td>Chi-Square</td>
<td>DF</td>
<td>Pr &gt; ChiSq</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1132.5331</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Score (Model-Based)</td>
<td>1346.7422</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Wald (Model-Based)</td>
<td>1179.2297</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

#### PANEL C: SUMMARY OF THE NUMBER OF EVENTS AND CENSORED STATES, 1984-2000

<table>
<thead>
<tr>
<th>Stratum (Sequence of events)</th>
<th>Down-grade estimation sample</th>
<th></th>
<th></th>
<th>Up-grade estimation sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>since the estimation period starts</td>
<td>Total</td>
<td>Event</td>
<td>Percent</td>
<td>Censored</td>
<td>Percent</td>
</tr>
<tr>
<td>1</td>
<td>1882</td>
<td>774</td>
<td>41.13</td>
<td>1108</td>
<td>58.87</td>
</tr>
<tr>
<td>2</td>
<td>1115</td>
<td>445</td>
<td>39.91</td>
<td>670</td>
<td>60.09</td>
</tr>
<tr>
<td>3</td>
<td>654</td>
<td>238</td>
<td>36.39</td>
<td>416</td>
<td>63.61</td>
</tr>
<tr>
<td>4</td>
<td>374</td>
<td>125</td>
<td>33.42</td>
<td>249</td>
<td>66.58</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>47</td>
<td>21.36</td>
<td>173</td>
<td>78.64</td>
</tr>
<tr>
<td>6</td>
<td>116</td>
<td>32</td>
<td>27.59</td>
<td>84</td>
<td>72.41</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>15</td>
<td>25</td>
<td>45</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>10</td>
<td>30.30</td>
<td>23</td>
<td>69.7</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>6</td>
<td>35.29</td>
<td>11</td>
<td>64.71</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>71.43</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>33.33</td>
<td>2</td>
<td>66.67</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4487</td>
<td>1698</td>
<td>37.84</td>
<td>2789</td>
<td>62.16</td>
</tr>
</tbody>
</table>

#### PANEL D: SUMMARY OF THE NUMBER OF EVENTS AND CENSORED STATES, 2001-2005

<table>
<thead>
<tr>
<th>Stratum (Sequence of event)</th>
<th>Downgrade holdout sample</th>
<th></th>
<th></th>
<th>Upgrade holdout sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>since the holdout period starts</td>
<td>Total</td>
<td>Event</td>
<td>Percent</td>
<td>Censored</td>
<td>Percent</td>
</tr>
<tr>
<td>1</td>
<td>1010</td>
<td>440</td>
<td>43.56</td>
<td>570</td>
<td>56.44</td>
</tr>
<tr>
<td>2</td>
<td>480</td>
<td>216</td>
<td>45</td>
<td>264</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>225</td>
<td>96</td>
<td>42.67</td>
<td>129</td>
<td>57.33</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>39</td>
<td>38.61</td>
<td>62</td>
<td>61.39</td>
</tr>
<tr>
<td>5</td>
<td>39</td>
<td>7</td>
<td>17.95</td>
<td>32</td>
<td>82.05</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>2</td>
<td>16.67</td>
<td>10</td>
<td>83.33</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1</td>
<td>20</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>Total</td>
<td>1872</td>
<td>801</td>
<td>42.79</td>
<td>1071</td>
<td>57.21</td>
</tr>
</tbody>
</table>
5.1.1. Overview

The discussion that follows focuses on the effects of the rating history variables, as their analysis is the purpose of this chapter. Consequently, there is no discussion of the macro-economic or industry sector control variables, except to note that an increase in output growth gap \( \text{RealGDPg\_actual\_minus\_potential} \) and stock market volatility \( \text{SP500\_annual\_SD} \) makes a downgrade more likely and an upgrade less likely.

In interpreting Panel A - Table 5.1, a negative coefficient reduces the hazard and therefore reduces the probability of a rating migration. The reported hazard ratios represent the relative change in the hazard for a one unit change in the independent variable. For example, in the model for downgrades, an increase in the length of the lag one rating run by one year reduces the likelihood of a downgrade by \( 1 - 0.923 \) or 7.7 percent.

The results in Panel A - Tables 5.1 are generally as hypothesised in chapter 2. The hazard of a rating change depends significantly upon several aspects of rating history as well as the current rating \( \text{start rating} \). The effects of longer durations at lag one and lag two \( \text{lag one, lag two} \), and a higher (better) \text{start rating} are the same for upgrades and downgrades, and the probability of a rating change in the current rating is reduced.

Consistent with Carty (1997), the results indicate that the better the current rating \( \text{start rating} \) the more likely the current rating state continues. The impact of the current rating is quite modest relative to the combined impact of the rating history variables. Indeed, it is modest in relation to some of the rating history variables considered individually. For example, in the model for downgrades a one unit (one
notch) increase in the current rating (start rating) reduces the downgrade hazards by 10.2 percent but a rating downgrade at lag one (dummy lag1 down) increases the hazard of a further downgrade by 156.9 percent.

There is some tendency for rating history to repeat itself. Consistent with Lando and Skodeberg (2002) and as hypothesised, longer lagged durations (lag one, lag two) increase the probability of the current rating continuing. In line with earlier studies such as Carty and Fons (1993), Altman and Kao (1992b), Kavvathas (2001), Bangia, et al. (2002), Lando and Skodeberg (2002), Hamilton and Cantor (2004), Figlewski et al. (2008), the results are consistent with rating momentum, particularly for downgrades.

The impact of lagged rating durations (lag one, lag two) is much smaller than the direction of lagged rating changes (dummy lag1 down, dummy lag2 down). For example, extending duration by a year at lags one and two reduces the probability of a downgrade by 7.7 percent and 6 percent respectively. In contrast, downgrades at lags one and two increase the hazard of a further downgrade by 156.9 percent and 23 percent respectively, which is consistent with momentum in rating migrations.\footnote{For upgrades a one-year increase in duration at lag one reduces the probability of an upgrade by 10.1 percent. Given a downgrade at lag one, the hazard of an upgrade is 70 percent of the hazard of cases where there was an upgrade at lag one. Therefore, an upgrade at lag one makes a further upgrade (1/0.7)-1 or 42.85 percent more likely.}

Furthermore, the duration of lag one and a downgrade at lag one have stronger effects on the hazard of a rating change than do the duration of lag two and a downgrade at lag two. For example, a downgrade at lag one (dummy lag1 down) makes a subsequent downgrade 156.9 percent more likely and a subsequent upgrade 30 percent less likely. But a downgrade at lag one two (dummy lag2 down) just increases...
the hazard of a further downgrade by 23 percent and does not affect the hazard of a further upgrade.

5.1.2. Comparison between downgrade and upgrade models

While the foregoing provides a general picture of the results, there are some differences between the upgrade and downgrade models. The downgrade model features a greater number of significant rating history variables, and downgrades are more severely impacted by rating history as evidenced by larger coefficient magnitudes. The models for upgrades and downgrades (Panel A - Table 5.1) have many rating history variables in common, but often their signs are reversed.

A downgrade at lag one (dummy lag1 down), and a higher downgrade volatility (rate prior down) make a subsequent downgrade more likely and a subsequent upgrade less likely. As hypothesised, a history of frequent rating downgrades is likely to be repeated. Increasing the downgrade volatility by one downgrade per year raises the hazard of a downgrade by 15.4 percent and reduces the hazard of an upgrade by 16.5 percent.

The incidence of being a fallen angel (number FA), and being within the investment/speculative rating boundary (dummy inv boundary/ dummy junk boundary) reduce the probability of a downgrade and increase the probability of an upgrade. Consistent with Mann et al., (2003), and Vazza et al., (2005), a surviving fallen angel is more likely to ascend the rating spectrum. The results show that the incidence of a fallen angel event makes a subsequent upgrade 22.6 percent more likely and a subsequent downgrade 33.8 percent less likely. In contrast to Johnson (2004), being in the investment-grade rating boundary (BBB+, BBB, BBB-) means that a firm is 28.7 percent less likely to fall to the lower rungs of rating scales, and 21.6 percent more
likely to ascend to higher investment grades. Being in the threshold of speculative-grade ratings (BB-, BB, BB+) has a similar effect on downgrades, but a stronger impact on upgrades; it increases the hazard of an upgrade by 61.1 percent. These findings are consistent with the hypothesis that issuers rated in the investment-grade boundary have considerable incentives to avoid losing their investment-grade status whereas those rated in the speculative-grade threshold have strong incentives to climb up the rating scales.

Additional variables in the downgrade model include rating age (age since first rated), the direction of the rating change at lag two state (dummy lag2 down), a break in the rating history (number NR), the incidence of being a rising star (number RS), and the occurrence of a substantial jump to higher rating grades (number big up). A period of being unrated (Number NR) reduces the probability of a downgrade by 38.9 percent. This result is consistent with issuers tending to withdraw from being rated to avoid a downward revision in rating. Consistent with Altman (1998), Figlewski et al. (2008), a greater age since first rated makes a subsequent downgrade more likely. However, the impact of rating age is the smallest of all the rating history variables in the down-grade model. The incidence of a rising star event (Number RS) and the occurrence of a substantial upgrade (Number big up) both increase the hazard of a downgrade. The effect of a substantial upgrade is much smaller than the effect of a rising star event as the latter increases the probability of a downgrade by 32 percent.

There is only one rating history variable in the upgrade model that is not shared with the downgrade model, and that is the original rating. The higher the original rating, the higher the probability of an upgrade, but the effect is the smallest of all the rating history variables in the upgrade model. An improvement of the original rating by one notch raises the hazard of an upgrade by just 3.3 percent.
In summary, while there are some differences between downgrades and upgrades it is clear that rating history has a strong impact on the hazard of a rating change. It is also clear that the rating history variables jointly, and in several cases individually, have a stronger impact than the current rating (start rating) on the hazard of a rating change. The results therefore suggest that rating changes are non-Markovian and history variables can be used in forecasting future rating changes. The question is how accurate are such forecasts?

5.2. The predictive accuracy of probability forecasts

The forecasting model as in equation (9) and (10) is based on the estimated baseline hazard function and the estimated coefficient vector from the in-sample analysis. The covariate vector from the holdout sample is then used to estimate the survival probabilities of holdout rating states at short-term forecast horizons ($t = 0.25, 0.5, 0.75, 1 \text{ year}$) and longer-term horizon ($t = 1, 2, 3, 4, 5 \text{ years}$).

The holdout sample changes through time. The forecast performance of the model through time depends on both the passage of time and changes in sample composition. Thus it is not possible to analyse the effect of time alone. However, we can see how the accuracy of rating migration forecasts varies through time for the firms that survived until the forecast time or beyond. This serves as the yardstick for comparison with the forecast performance of the time-varying covariate (TVC) models developed in chapter 6.
### Table 5-2

**BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TIME-FIXED DOWNGRADE MODEL**

<table>
<thead>
<tr>
<th>Forecast horizon $t$ (year)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>Naïve Brier score*</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Benchmark Brier score##</td>
<td>0.3670</td>
<td>0.2929</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2448</td>
<td>0.2199</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1222</td>
<td>0.0730</td>
</tr>
<tr>
<td>Forecast model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of forecasts###</td>
<td>1872</td>
<td>1348</td>
</tr>
<tr>
<td>Mean probability of outcome</td>
<td>0.5721</td>
<td>0.6736</td>
</tr>
<tr>
<td>Mean probability of forecast</td>
<td>0.9414</td>
<td>0.9163</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.3729</td>
<td>0.275</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2448</td>
<td>0.2199</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1281</td>
<td>0.0551</td>
</tr>
<tr>
<td>Forecast variance (scatter)</td>
<td>0.003</td>
<td>0.0048</td>
</tr>
<tr>
<td>Reliability-in-the-large (bias square)</td>
<td>0.1364</td>
<td>0.0589</td>
</tr>
<tr>
<td>2*Forecast-Outcome-Covariance</td>
<td>0.0114</td>
<td>0.0086</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0233</td>
<td>0.0196</td>
</tr>
</tbody>
</table>

* The naïve Brier score is constructed by setting the probability forecast at time $t$ in equation (11) equal to a random forecast of 0.5.

## The Brier score of the benchmark was constructed by setting the probability forecast at time $t$ in equation (11) equal to the proportion of rating states surviving beyond time $t$ in the estimation sample.

### The sample size of forecasts at one-year forecast horizon in Panel A and Panel B differs for the reason explained in chapter 4.
Figure 5-1

BRIER SCORE COVARIANCE DECOMPOSITIONS OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TIME-FIXED DOWNGRADE MODEL

Panel A: Short-term forecasts

Covariance graph_Down-states, time=0.5

Covariance graph_Down-states, time=0.75

Covariance graph_Down-states, time=1

Mean survival

Mean survival

Mean survival

Outcome index

Outcome index

Outcome index

Estimated probability of survival

Estimated probability of survival

Estimated probability of survival

Migration frequency histogram

Migration frequency histogram

Migration frequency histogram

Survival frequency histogram

Survival frequency histogram

Survival frequency histogram

Brier score 0.275, Scatter 0.0048, Bias 0.2427, Slope 0.0196

Brier score 0.2203, Scatter 0.0093, Bias 0.1572, Slope 0.0221

Brier score 0.1908, Scatter 0.0135, Bias 0.0954, Slope 0.0157

Panel B: Longer-term forecasts

Covariance graph_Down-states, time=1

Covariance graph_Down-states, time=2

Covariance graph_Down-states, time=3

Mean survival outcome index

Mean survival probabilistic forecast

Regression line for forecasts on outcome indexes

Regression line for forecasts on outcome indexes

Regression line for forecasts on outcome indexes

Migration frequency histogram

Survival frequency histogram

Survival frequency histogram

Survival frequency histogram

Brier score 0.3001, Scatter 0.0237, Bias 0.2642, Slope 0.0781

Brier score 0.1714, Scatter 0.0286, Bias -0.08, Slope 0.0458

Brier score 0.2109, Scatter 0.0369, Bias -0.3253, Slope 0.0396
The Brier scores (Brier, 1950) of forecasts generated at the various forecast horizons are calculated as in equation (11). As explained in chapter 4, using the covariance decomposition (Yates, 1982), the Brier score can be broken down into skill components, which suggest reasons for discrepancies between the observed outcome and probability forecast at a given time horizon.

Lower Brier scores and lower skill component scores imply better forecast ability. The skill components of the estimated models are then assessed in reference to the “comparable” skill components of benchmark models. The following analyses the forecast performance in terms of the skill components (bias, slope, scatter). The focus is on the sources of forecast errors and making suggestions to improve the forecast performance of the models.

5.2.1. Downgrade models

5.2.1.1. Overview

The Brier scores of forecasts generated by the downgrade models are summarised in Panel A (short-term horizon) and Panel B (longer-term forecast horizons) Table 5-2. Figure 5-1 depicts the covariance graphs of survival forecasts generated at short-term horizon (Panel A, forecast time $t = 0.5, 0.75$ and 1 year) and longer-term horizon (Panel B, forecast time $t = 1, 2, 3$ years).

In the aggregate, the downgrade model does not consistently outperform either the benchmark or the naïve forecasts. Within the one-year window, however, the model has an edge at the three quarters and one-year forecast horizons (Panel A Table 5-2).

Overall, the survival forecasts have a large bias. As a consequence of this poor calibration, the model overestimates survival within the one-year window (Panel A and B, Figure 5-1). A striking feature is that most of the estimates within the one-year
forecast horizon were placed in the probability bins above the 80 percent probability of survival. The model seldom offers pessimistic estimates of survival. The strong tendency to place estimates in the optimistic probability categories, substantially higher than the relative survival frequency, indicates that the model fails to capture the short duration of down states. The scatter is small, but discrimination is poor. The regression line is almost flat as the mean survival forecasts for surviving states and non-surviving states are quite close (Panel A Figure 5-1).

The composition of the holdout sample dramatically changes at yearly forecast intervals (relative to quarterly forecast intervals) and the credit profile of issuers in the holdout sample improves considerably as time progresses. This is reflected in the substantially higher survival base rates as the forecast horizon extends beyond year one. However, the model fails to capture this, and instead uses optimistic categories far less often at the two and three-year forecast horizons. For instance, only 43.64 percent of two-year estimates and 14.29 percent of three-year estimates were placed in the categories above 80 percent. On the other hand, the model offers pessimistic forecasts more frequently (relative to one-year estimates). 10.46 percent of two-year forecasts and 27.11 percent of three-year forecasts were assigned to the categories below 50 percent. The tendency to place a high proportion of survival forecasts in the categories lower than the relative survival frequency translates into negative bias at year two and thereafter. In other words, the model overestimates downgrades and underestimates survival probabilities as the forecast horizon extends beyond one year (Panel B, Figure 5-1).

5.2.1.2. Sources of forecast errors and implications

The following discussion considers three factors that contribute to the large bias and poor discrimination of the survival forecasts.
First, the estimation period covers a substantially longer period than the holdout period, smoothing out the cyclical effects of macro-economic change. Additionally, the rating dynamics in the estimation period are not representative of rating changes in the holdout period. The common notion is that corporate credit quality changes over time (Carty and Fons, 1993). Nickell et al. (2000) further indicate that the migration dynamics vary over the business cycle and the occurrences of downgrades increase as market conditions deteriorate. The period 2001-2005 was markedly different from the period 1984-2000. As depicted in Figure 4-7 and Figure 4-8, it was characterised by dramatic deterioration in the credit market, and prolonged downturns in the economy. As shown in Figures 4-11 and 4-12, the holdout period contained a substantially greater proportion of down-grades in the low rating categories than did the estimation period. It is also clear that down states in the holdout sample are generally short-lived and mass at durations of less than one year (Figure 4-15). The mean survival time of down states in the holdout period was 0.514 year, less than one third of the mean survival time of down states in the estimation sample (Table 4-8). In other words, it took substantially less time for a typical downgrade to occur in the holdout period than in the estimation period. Consequently, the short-term survival base rates of holdout states are lower than the corresponding base rates of estimation states. It is, therefore, not surprising that the model forecasts based on the “average” migration experience in the long period, 1984-2000, fail to capture the volatile downgrade dynamics of the subsequent volatile period.

Second, the estimation model does not allow for variation in firm-specific and macro-economic covariates, as time passes. In particular, the static model fails to capture the evolution of the macro-economic environment and the conditions that led to a rapid
deterioration in credit quality in the holdout sample. This suggests the need to develop a dynamic model that includes time-varying covariates.

Third, a common conjecture is that credit rating agencies generally focus more resources on quantifying deteriorations in the credit profile of an issuer than they do on assessing the improvement in performance. Downgrades tend to be swift whereas upgrades lag the improvement in credit quality. As suggested by Amato and Furfine (2003), credit rating agencies exhibit a propensity to overreact to the prevailing conditions when they revise ratings. They tend to be over-optimistic when the economy performs well and display excessive pessimism in downturns. These factors will have contributed to an acceleration of downgrades during the poor macro-economic conditions of the holdout period and this acceleration was not captured by the model.

5.2.2. Upgrade models

5.2.2.1. Overview

The Brier score decompositions of upgrade forecasts are summarised in Panel A (short-term forecasts) and Panel B (longer-term forecasts) of Table 5-3 below. For the purpose of illustration, Figure 5-2, which follows Table 5-3, depicts the covariance graphs of survival forecasts generated by the upgrade model at a short-term horizon (Panel A, forecast time $t=0.5, 0.75$ and $1$ year) and longer-term horizon (Panel B, forecast time $t=1, 2,$ and $3$ years).
Table 5-3

<table>
<thead>
<tr>
<th>Forecast horizon $t$ (year)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25 0.5 0.75 1</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Naïve Brier score#</td>
<td>0.25 0.25 0.25 0.25</td>
<td>0.25 0.25 0.25 0.25 0.25</td>
</tr>
<tr>
<td>Benchmark Brier score##</td>
<td>0.129 0.1493 0.1671 0.1772</td>
<td>0.1178 0.1749 0.1225 0.0698 0.0736</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.1137 0.13 0.1442 0.1546</td>
<td>0.1137 0.1681 0.1224 0.0694 0.071</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0153 0.0193 0.0229 0.0226</td>
<td>0.0041 0.0068 0.0001 0.0004 0.0026</td>
</tr>
</tbody>
</table>

**Forecast model**

| Number of forecasts### | 1872 1348 1059 832       | 1872 669 273 80 13 |
| Mean probability of outcome | 0.8691 0.8464 0.8253 0.8089 | 0.8691 0.7862 0.8571 0.925 0.9231 |
| Mean probability of forecast | 0.9651 0.9454 0.9267 0.8991 | 0.8863 0.778 0.6506 0.5505 0.3823 |
| Brier score              | 0.1221 0.1393 0.1527 0.1582 | 0.1124 0.1593 0.1704 0.243 0.3602 |
| Outcome index variance   | 0.1137 0.13 0.1442 0.1546   | 0.1137 0.1681 0.1224 0.0694 0.071  |
| Skill components         | 0.0084 0.0093 0.0085 0.0036 | -0.0013 -0.0088 0.048 0.1736 0.2892 |
| Forecast variance (scatter) | 0.0007 0.0008 0.0012 0.0022 | 0.0036 0.0116 0.0227 0.026 0.0108 |
| Reliability-in-the-large (bias square) | 0.0092 0.0098 0.0103 0.0081 | 0.0003 0.0001 0.0426 0.1403 0.2924 |
| 2^Forecast-Outcome-Covariance | 0.0015 0.0013 0.003 0.0067 | 0.0053 0.0204 0.0175 -0.0073 0.014 |
| Slope                    | 0.0066 0.0050 0.0104 0.0217 | 0.0233 0.0607 0.0715 -0.0526 0.0986 |

# The naïve Brier score is constructed by setting the probability forecast at time $t$ in equation (11) equal to a random forecast of 0.5.

## The Brier score of the benchmark was constructed by setting the probability forecast at time $t$ in equation (11) equal to the proportion of rating states surviving beyond time $t$ in the estimation sample.

### The sample size of forecasts at one-year forecast horizon in Panel A and Panel B differs for the reason explained in chapter 4.
Figure 5-2

BRIER SCORE COVARIANCE DECOMPOSITIONS OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TIME-FIXED UPGRADE MODEL

Panel A: Short-term forecasts

Covariance graph_Up-states, time=0.5

Brier score 0.1393, Scatter 0.0008, Bias 0.099, Slope 0.004895

Covariance graph_Up-states, time=0.75

Brier score 0.1527, Scatter 0.0012, Bias 0.1015, Slope 0.010289

Covariance graph_Up-states, time=1

Brier score 0.1582, Scatter 0.0022, Bias 0.09, Slope 0.021665

Panel B: Longer-term forecasts

Covariance graph_Up-states, time=0.5

Brier score 0.1124, Scatter 0.0036, Bias 0.017, Slope 0.02335

Covariance graph_Up-states, time=0.75

Brier score 0.1593, Scatter 0.0116, Bias -0.01, Slope 0.060745

Covariance graph_Up-states, time=1

Brier score 0.1704, Scatter 0.0227, Bias -0.2064, Slope 0.071351

Brier score 0.1582, Scatter 0.0265, Bias 0.09, Slope 0.021665
Overall, the upgrade model exhibits some predictive accuracy. Within the two year forecast horizon, it outperforms both the benchmark and naïve forecasts by a small margin, owing to a minimal bias and negligible scatter. The upgrade model differs markedly from the downgrade model in that it offers well-calibrated short-term forecasts, but it has relatively low discrimination.

Similar to the downgrade model, the upgrade model does well in removing irrelevant information (scatter) but at the expense of failing to incorporate some information that is relevant to the event occurrence (slope). The model performs poorly in discriminating rating states that survived from states that migrated (upgraded).

The covariance graphs in Panel A of Figure 5-2 show that the upgrade model generally assigns short-term forecasts to the most optimistic categories (above 80 percent). In contrast to downgrades, this does not result in a large bias. This is because up states have a higher survival base rate and longer average survival time than down states.

A striking feature in Panel B of Figure 5-2 is that the mean estimate line and the base rate line cross almost exactly on the 45-degree line at one and two-year forecast horizons. The model is thus very well calibrated at the one and two-year forecast horizons. At these horizons, the model achieves negative skill components primarily because of negligible bias and minimal scatter (Panel B Table 5-3). However, the slope shows relatively little discrimination at both short-term and longer-term horizons.

At horizons from three years onwards, the performance of the model deteriorates sharply and fails to outperform the benchmark model. In part this is due to increasing

---

35 A negative skill component occurs when the sum of bias square and scatter is smaller than two times covariance.
negative bias as the model increasingly makes forecasts that lie in the lower probability bins.

5.2.2.2. Sources of forecast errors and implications

Why does the upgrade model perform better than the downgrade model? Why does the performance of the upgrade model deteriorate sharply after the three-year horizon? The following discussion addresses these questions.

As discussed chapter 4, the rating dynamics in the holdout period 2001-2005 differ from the dynamics in the 1984-2000 estimation period. The proportion of upgrades declined whereas the proportion of downgrades soared. The change for upgrades was not as dramatic as for downgrades.

In times of market turbulence, as in 2002-2003, issuers with a deteriorating credit profile (downgrade candidates) are naturally under the close scrutiny of rating agencies and become targets for downward revisions. On the other hand, obligors with an improving credit profile (upgrade candidates) may be less frequently reviewed by credit rating agencies. In addition, credit rating agencies tend to be reluctant to revise rating upwards during economic downturns. As a result, upgrades in the holdout sample tend to be less volatile than downgrades and are therefore easier to forecast.

As the forecast horizon extends and firms of high migration (upgrade) “risk” leave the holdout sample, the model fails to reflect the resulting improvement in survival base rates over time. Consequently, performance deteriorates sharply from year three onwards. The information contained in rating history also becomes increasingly stale as the rating duration lengthens. Allowing for interactions between the time spent in a rating grade and rating history might improve the forecasts.
5.3. Conclusion

This chapter develops stratified Cox’s proportional hazard models to address two issues. Firstly, to examine non-Markovian behaviour in rating dynamics and secondly, to assess the predictive accuracy of time-varying probability forecasts. Two proportional hazard models for rating upgrades and downgrades during the period 1984-2000 were developed. The significance and sign of rating history variables vary between downgrades and upgrades and the downgrade model contains more significant variables.

As hypothesised, there is statistically significant dependence of the migration hazard on rating history variables and the current rating (start rating). Importantly, the impact of the current rating (start rating) is modest compared to the effect of the rating history variables. For example, the direction of a rating change at lag one (dummy lag1 down) is much more important than the current rating. These results clearly suggest that migration hazards depend on rating history, and reinforce the literature that calls into question the Markov assumption common in the analysis of rating migrations.

For both rating upgrades and downgrades, the longer the length of lag one and lag two rating states, the more likely a rating continues in its current grade. In the downgrade model, experiencing a downgrade at lag one (dummy lag1 down) makes a downgrade more likely and in the upgrade model it makes an upgrade less likely. These results are consistent with momentum in rating grades.

The Yates’s covariance decomposition of Brier score was used to assess the predictive accuracy of probability forecasts generated by the downgrade and upgrade models at quarterly intervals (t=0.25, 0.5, 0.75, 1 year) and yearly intervals (t=1, 2, 3,
4, 5 years) over the period 2001-2005. The forecast performance is affected by both time and the changing sample composition.

In aggregate the performance of the models is disappointing. The models have some predictive power, but their ability to discriminate rating states that survived from states that migrated is generally poor. The upgrade model generally performs better than the downgrade model. However, both upgrade and downgrade models display bias. They have a strong tendency to overestimate the probability of survival at short horizons and underestimate the probability of survival at longer horizons.

An important part of the possible explanation for the poor performance of the models is that following the estimation period economic conditions deteriorated. Hence rating dynamics changed dramatically in the holdout period. The assessment of forecast accuracy suggests two directions for improving the model. One is to add time-varying covariates to the model and in so doing allow for changing credit conditions through time. This is the subject of subsequent chapters. Another way of trying to capture changing conditions would be to continually update the model by using a moving window and/or recalibrating the model as changes in rates of migration become evident. This is a topic for future research.
Chapter 6

Rating migrations: The effect of rating history and time

This chapter extends the work developed in chapter 5 and estimates two stratified dynamic Cox’s hazard models with time-varying covariates (TVC), namely the TVC base model and the TVC extended model, for each down state and up state in the estimation sample.

The TVC base model incorporates the same set of firm-specific (rating history) and macro-economic covariates as the Cox’s proportional hazard model (time-fixed model) developed in chapter 5. The difference between the two models lies in the construction of macro-economic covariates. In the time-fixed model lagged values of macro-economic covariates are measured at the beginning of each rating state. The time-fixed model therefore controls for the economic environment prevailing at the start of each rating state. On the other hand, the TVC base model includes both time-fixed and time-varying macro-economic covariates. Lagged values of time-varying covariates are updated quarterly, and the estimation process incorporates the lagged values measured prior to, or at, each event time. The TVC base model, therefore, takes into account both the state of the economic cycle at the commencement of each rating state and the evolution of macro-economic conditions over the rating duration.

The relevant question is whether the impact of rating history on migration hazards is intact in the presence of time-varying macro-economic covariates? The results of the TVC base model indicate that the dynamic estimation framework does not alter the effects of rating history on migration hazards.
The TVC extended model extends the TVC base model by incorporating both time-varying macro-economic covariates and the interaction terms between rating history and time. In the estimation process, the interaction terms are constructed as time-varying covariates and are updated whenever an event occurs. The TVC extended model therefore captures not only the economic changes over rating durations but also accounts for the interactions between rating history and the passage of survival time.

The relevant question is whether time matters and how the interactions with time alter the main effects of rating history on migration hazards? The results of the TVC extended model show that there is evidence of significant interactions between the time spent in a rating grade and the main effect variables. The impact of the rating history variables diminishes as the time spent in the current rating extends.

The stratified TVC base and stratified TVC extended models given by equation (3) were estimated for the upgrades and downgrades during the estimate period 1984-2000. As in chapter 5, in the interest of a parsimonious model and compact presentation of the results, a backward stepwise estimation procedure was employed. Variables were retained in the models according to the log-likelihood ratio test, at the 10 percent level or better.

The following discussion focuses on the comparative results of the TVC base model relative to the time-fixed model, and the TVC extended model relative to the TVC base model for the period 1984-2000.

6.1. Stratified TVC base model versus stratified time-fixed model

The estimation results of the TVC base models and the time-fixed models for downgrades and upgrades are presented in Panel A - Table 6.1. Panel B provides statistics on the fit of the models.
### Table 6.1

**STRATIFIED TIME VARYING COVARIATE (TVC) BASE MODEL AND TIME FIXED MODEL**  
**PANEL A: COEFFICIENT ESTIMATES (BACKWARD SELECTION), 1984-2000**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rating history</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age_since_first_rated</td>
<td>0.03777***</td>
<td>0.01126</td>
<td>1.038</td>
<td>0.03691***</td>
<td>0.01092</td>
<td>1.038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original_rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start_rating</td>
<td>-0.10267***</td>
<td>0.00802</td>
<td>0.902</td>
<td>-0.32779***</td>
<td>0.06977</td>
<td>0.721</td>
<td>-0.31063***</td>
<td>0.08471</td>
<td>0.733</td>
<td>-0.07707***</td>
<td>0.01948</td>
<td>0.926</td>
<td>0.9107***</td>
<td>0.06606</td>
<td>2.486</td>
</tr>
<tr>
<td>Dummy_inv_boundary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag_one</td>
<td>-0.06212***</td>
<td>0.01967</td>
<td>0.94</td>
<td>0.21756***</td>
<td>0.065</td>
<td>1.243</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy_lag1_down</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag_two</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy_lag2_down</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate_prior_change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number_NR</td>
<td>-0.51913***</td>
<td>0.11748</td>
<td>0.595</td>
<td>-0.41478***</td>
<td>0.08669</td>
<td>0.66</td>
<td>0.28122***</td>
<td>0.1006</td>
<td>1.325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number_FA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number_RS</td>
<td>0.17623***</td>
<td>0.04146</td>
<td>1.193</td>
<td>-0.13395***</td>
<td>0.03032</td>
<td>0.875</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number_big_down</td>
<td>0.08793*</td>
<td>0.05146</td>
<td>1.092</td>
<td>0.21222*</td>
<td>0.11913</td>
<td>1.236</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number_big_up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Macroeconomic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy_recession</td>
<td>0.21222*</td>
<td>0.11913</td>
<td>1.236</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFNAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RealGDPg_actual_minus_potential</td>
<td>0.23056*</td>
<td>0.1283</td>
<td>1.257</td>
<td>-0.22581***</td>
<td>0.06644</td>
<td>0.798</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industrial_production_change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP500_quarterly_return</td>
<td>-0.01675*</td>
<td>0.00903</td>
<td>0.983</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP500_annual_SD</td>
<td>0.17623***</td>
<td>0.04146</td>
<td>1.193</td>
<td>-0.13395***</td>
<td>0.03032</td>
<td>0.875</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term_structure_slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Industry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerospace/automotive/capital goods/metal</td>
<td>-0.13803*</td>
<td>0.07283</td>
<td>0.871</td>
<td>-0.39024***</td>
<td>0.12142</td>
<td>0.677</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer / service sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy and natural resources</td>
<td>-0.39024***</td>
<td>0.12142</td>
<td>0.677</td>
<td>-0.43533***</td>
<td>0.12814</td>
<td>0.784</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telecommunications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure time / media</td>
<td>-0.432***</td>
<td>0.14577</td>
<td>0.648</td>
<td>-0.25076***</td>
<td>0.88672</td>
<td>0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Estate</td>
<td>-0.24597***</td>
<td>0.0737</td>
<td>0.782</td>
<td>-0.24597***</td>
<td>0.07398</td>
<td>0.772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility</td>
<td>-0.21071*</td>
<td>0.11833</td>
<td>0.81</td>
<td>-0.21071*</td>
<td>0.11833</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest and building products/homebuilders</td>
<td>-0.18775</td>
<td>0.12105</td>
<td>0.829</td>
<td>-0.18775</td>
<td>0.12105</td>
<td>0.829</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care / chemicals</td>
<td>-0.24597***</td>
<td>0.0737</td>
<td>0.782</td>
<td>-0.24597***</td>
<td>0.07398</td>
<td>0.772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transportation</td>
<td>-0.24597***</td>
<td>0.0737</td>
<td>0.782</td>
<td>-0.24597***</td>
<td>0.07398</td>
<td>0.772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High technology/computers/office equipment</td>
<td>-0.25115*</td>
<td>0.14609</td>
<td>0.778</td>
<td>-0.25115*</td>
<td>0.14609</td>
<td>0.778</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** P-value ≤ 1%, ** 1% < P-value ≤ 5%, * 5% < P-value ≤ 10% based on Wald chi-square tests
### Table 6.1 (Continued)

**STRATIFIED TIME VARYING COVARIATE (TVC) BASE MODEL AND TIME FIXED MODEL**

**PANEL B: TESTING GLOBAL NULL HYPOTHESIS $\beta=0$, 1984-2000**

<table>
<thead>
<tr>
<th></th>
<th>-2LogL</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
<th>-2LogL</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
<th>-2LogL</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
<th>-2LogL</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>1016.93</td>
<td>23</td>
<td>&lt;.0001</td>
<td>1132.53</td>
<td>23</td>
<td>&lt;.0001</td>
<td>450.97</td>
<td>21</td>
<td>&lt;.0001</td>
<td>463.06</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score (Model-Based)</td>
<td>1157.12</td>
<td>23</td>
<td>&lt;.0001</td>
<td>1346.74</td>
<td>23</td>
<td>&lt;.0001</td>
<td>444.84</td>
<td>21</td>
<td>&lt;.0001</td>
<td>454.57</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald (Model-Based)</td>
<td>1036.42</td>
<td>23</td>
<td>&lt;.0001</td>
<td>1179.23</td>
<td>23</td>
<td>&lt;.0001</td>
<td>422.31</td>
<td>21</td>
<td>&lt;.0001</td>
<td>431.67</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**PANEL C: SUMMARY OF THE NUMBER OF EVENTS AND CENSORED STATES, 1984-2000**

<table>
<thead>
<tr>
<th>Stratum (Sequence of events)</th>
<th>Total</th>
<th>Event</th>
<th>Percent Event</th>
<th>Censored</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1882</td>
<td>774</td>
<td>41.13</td>
<td>1108</td>
<td>58.87</td>
</tr>
<tr>
<td>2</td>
<td>1115</td>
<td>445</td>
<td>39.91</td>
<td>670</td>
<td>60.09</td>
</tr>
<tr>
<td>3</td>
<td>654</td>
<td>238</td>
<td>36.39</td>
<td>416</td>
<td>63.61</td>
</tr>
<tr>
<td>4</td>
<td>374</td>
<td>125</td>
<td>33.42</td>
<td>249</td>
<td>66.58</td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>47</td>
<td>21.36</td>
<td>173</td>
<td>78.64</td>
</tr>
<tr>
<td>6</td>
<td>116</td>
<td>32</td>
<td>27.59</td>
<td>84</td>
<td>72.41</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>15</td>
<td>25</td>
<td>45</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>10</td>
<td>30.3</td>
<td>23</td>
<td>69.7</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>6</td>
<td>35.29</td>
<td>11</td>
<td>64.71</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>71.43</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>33.33</td>
<td>2</td>
<td>66.67</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4487</td>
<td>1698</td>
<td>37.84</td>
<td>2789</td>
<td>62.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Event</th>
<th>Percent Event</th>
<th>Censored</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downgrade TVC base model</td>
<td>453</td>
<td>24.07</td>
<td>1429</td>
<td>75.93</td>
</tr>
<tr>
<td>Downgrade time fixed model</td>
<td>296</td>
<td>26.55</td>
<td>819</td>
<td>73.45</td>
</tr>
<tr>
<td>Upgrade TVC base model</td>
<td>175</td>
<td>26.76</td>
<td>479</td>
<td>73.24</td>
</tr>
<tr>
<td>Up-grade time fixed model</td>
<td>110</td>
<td>29.41</td>
<td>264</td>
<td>70.59</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td>33.64</td>
<td>146</td>
<td>66.36</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>28.45</td>
<td>83</td>
<td>71.55</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>33.33</td>
<td>40</td>
<td>66.67</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>24.24</td>
<td>25</td>
<td>75.76</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.76</td>
<td>15</td>
<td>88.24</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>71.43</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>50</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>1175</td>
<td>26.19</td>
<td>3312</td>
<td>73.81</td>
</tr>
</tbody>
</table>
The results of the TVC base model are generally as hypothesised and consistent with the results of the time-fixed models. The TVC base downgrade/upgrade model and the respective time-fixed model feature the same set of significant rating history variables. There is little variation in the individual effects of rating history across the two models. The results for the TVC base models therefore reinforce the key findings of the time-fixed models. Rating history variables are significant determinants of rating migration hazards. Relative to rating history, the current rating (start rating) has a small effect. There is strong evidence that rating history repeats. Rating dynamics exhibit serial correlations in the direction of changes (dummy lag1 down), and duration dependence (lag one, lag two). The impact of rating momentum (dummy lag1 down) is particularly acute for downgrades. A longer lagged rating duration (lag one, lag two) increases the probability that a state will remain in its current rating grade.

The effects of some aspects of rating history vary between downgrades and upgrades. For instance, a history of frequent downgrades (rate prior down) makes a downgrade more likely and an upgrade less probable. The occurrence of a fallen angel event (Number FA), and being close to the investment-grade or speculative-grade rating threshold (Dummy inv boundary/Dummy junk boundary) reduce the downgrade risk but increase the upgrade probability. In the aggregate, downgrades are more affected by rating history than upgrades, as evidenced by the presence of additional significant rating history variables (age since first rated, dummy lag2 down, Number NR, Number RS, Number big up).

While the above summarises the similarities, there are some differences between the TVC base model and the time-fixed model. The distinguishing features lie in the presence of significant macro-economic covariates.
For downgrades, compared with the time-fixed model, the TVC base model includes a greater number of significant macro-economic covariates, although CFNAI is no longer significant. The output growth gap (RealGDPg_actual_minus_potential), SP500 annual SD and term structure slope retain the same sign. An increase in output growth gap, a higher volatility in stock market performance, and a decline in term structure slope raise the downgrade hazard, though the magnitude of each variable varies between the two models. This is not surprising as downgrades are more frequently observed at times of volatile stock market performance and tightened credit conditions. Notable changes in the TVC base downgrade model are the significant presence of the recession index (dummy recession), industrial production change, and SP500 quarterly return. A decrease in industrial production change and SP500 quarterly return, and being in a recession (dummy recession) make a downgrade more likely. These results are intuitively appealing, as an economic recession or a contraction, characterised by a decline in stock market returns, a decrease in industrial activities and a flatter term structure, tends to produce more downgrades.

For upgrades, relative to the time-fixed model, the TVC base model features fewer significant macro-economic covariates. The recession index (dummy recession) and SP500 quarter return have a negative impact in the time-fixed model; however, these variables are not significant in the TVC base model. Only output growth gap (RealGDPg_actual_minus_potential ) and stock market volatility (SP500 annual SD) remain significant. As in the time-fixed model, a higher volatility in the stock market performance makes an upgrade less likely. Of particular interest, RealGDPg_actual_minus_potential changes its sign. A one percent increase in output growth gap makes an upgrade 25.2 percent less likely in the time-fixed model but 38.4 percent more likely in the TVC model.
In the aggregate, the individual impacts of rating history variables remain intact in the presence of time-varying macro-economic covariates. This result is consistent with Figlewski et al. (2008). The information provided by time-varying macro-economic variables is “incremental” compared to the information contained in rating history variables.

The modest impact of the time-varying macro-economic covariates can be attributed to the way these variables were constructed and the short-lived nature of rating states. Macro-economic variables, except the recession index (dummy recession), were constructed as either an exponentially weighted average of lagged observations computed quarterly over a six-quarter window or a three-month moving average (CFNAI). As rating states, particularly down states, are short lived, the macro-economic covariates measured at the commencement of each rating state and prior to each migration time are unlikely to be very different.

6.2 Stratified TVC extended model versus stratified TVC base model

The estimation results of the TVC base models, and the TVC models extended by the time interaction variables, are given for downgrades and upgrades in Panel A and Panel B - Table 6.2. The results in Panel A - Table 6.2 are generally as hypothesised. Rating history covariates are key determinants of rating migration hazards. It is also clear that there are significant interactions with time for most rating history variables. The evidence of interactions was more frequently observed for downgrades than for upgrades. The interaction effects are generally of the opposite sign to the main effects and this is consistent with decay in the impact of rating history the longer a rating remains unchanged.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Downgrade TVC base model</th>
<th>Downgrade TVC extended model</th>
<th>Upgrade TVC base model</th>
<th>Upgrade TVC extended model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter estimate</td>
<td>Standard error</td>
<td>Hazard ratio</td>
<td>Parameter estimate</td>
</tr>
<tr>
<td><strong>Rating history</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age_since_first_rated</td>
<td>0.03777***</td>
<td>0.01126</td>
<td>1.038</td>
<td>0.03777***</td>
</tr>
<tr>
<td>Original_rating</td>
<td>0.033***</td>
<td>0.01099</td>
<td>1.034</td>
<td>0.033***</td>
</tr>
<tr>
<td>Start_rating</td>
<td>-0.10267***</td>
<td>0.00802</td>
<td>0.902</td>
<td>-0.10267***</td>
</tr>
<tr>
<td>Dummy_inv_boundary</td>
<td>-0.32779***</td>
<td>0.06977</td>
<td>0.721</td>
<td>-0.32779***</td>
</tr>
<tr>
<td>Dummy_junk_boundary</td>
<td>-0.31063***</td>
<td>0.08471</td>
<td>0.733</td>
<td>-0.31063***</td>
</tr>
<tr>
<td>Lag_one</td>
<td>-0.10267***</td>
<td>0.00802</td>
<td>0.902</td>
<td>-0.10267***</td>
</tr>
<tr>
<td>Dummy_lag1_down</td>
<td>0.9107***</td>
<td>0.06606</td>
<td>2.486</td>
<td>1.40837***</td>
</tr>
<tr>
<td>Lag_two</td>
<td>-0.06212***</td>
<td>0.01967</td>
<td>0.94</td>
<td>-0.0502**</td>
</tr>
<tr>
<td>Rate_prior_change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate_prior_down</td>
<td>0.11973*</td>
<td>0.06836</td>
<td>1.127</td>
<td>-0.20078**</td>
</tr>
<tr>
<td>Number_NR</td>
<td>-0.51913***</td>
<td>0.11748</td>
<td>0.565</td>
<td>-0.79337***</td>
</tr>
<tr>
<td>Number_FA</td>
<td>-0.41478***</td>
<td>0.08669</td>
<td>0.66</td>
<td>-0.4105**</td>
</tr>
<tr>
<td>Number_RS</td>
<td>0.28122**</td>
<td>0.1006</td>
<td>1.325</td>
<td>0.31811***</td>
</tr>
<tr>
<td>Number_big_down</td>
<td>0.12183</td>
<td>0.0782</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Number_big_up</td>
<td>0.08793*</td>
<td>0.05146</td>
<td>1.092</td>
<td>0.09382*</td>
</tr>
<tr>
<td>Original_rating_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Start_rating_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Inv_boundary_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Junk_boundary_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Lag_one</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Lag1_down</td>
<td>-0.26154</td>
<td>0.03699</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Lag_two</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Lag2_down</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>NR_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>FA_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>RS_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Big_down</td>
<td>-0.10008</td>
<td>0.04091</td>
<td>0.905</td>
<td></td>
</tr>
<tr>
<td>Big_up_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Log_age_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Rate_change_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Rate_down_time</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td><strong>Macroeconomic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy_recession</td>
<td>0.21222*</td>
<td>0.11913</td>
<td>1.236</td>
<td>0.20414*</td>
</tr>
<tr>
<td>CFNAI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RealGDPg_actual_minus_potential</td>
<td>0.23056*</td>
<td>0.1283</td>
<td>1.259</td>
<td>0.22772*</td>
</tr>
<tr>
<td>Industrial_production_change</td>
<td>-0.22581***</td>
<td>0.06644</td>
<td>0.798</td>
<td>-0.25943***</td>
</tr>
<tr>
<td>SP500_quarterly_return</td>
<td>-0.01675*</td>
<td>0.00903</td>
<td>0.983</td>
<td></td>
</tr>
<tr>
<td>SP500_annual_SD</td>
<td>0.17623*</td>
<td>0.04146</td>
<td>1.193</td>
<td>0.18443*</td>
</tr>
<tr>
<td>Term_structure_slope</td>
<td>-0.13395**</td>
<td>0.03032</td>
<td>0.875</td>
<td>-0.09746**</td>
</tr>
</tbody>
</table>
## Table 6.2 (Continued)

### STRATIFIED TIME VARYING COVARIATE MODEL

#### PANEL A: COEFFICIENT ESTIMATES (BACKWARD SELECTION), 1984-2000

<table>
<thead>
<tr>
<th>Variables</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Hazard ratio</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Hazard ratio</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Hazard ratio</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indusry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerospace/automotive/capital goods/metal</td>
<td>-0.13803**</td>
<td>0.07283</td>
<td>0.871</td>
<td>-0.53407***</td>
<td>0.12174</td>
<td>0.586</td>
<td>-0.56606***</td>
<td>0.11902</td>
<td>0.568</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer / service sector</td>
<td>0.17217**</td>
<td>0.07097</td>
<td>1.188</td>
<td>-0.69284***</td>
<td>0.12699</td>
<td>0.5</td>
<td>-0.69618***</td>
<td>0.12407</td>
<td>0.498</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy and natural resources</td>
<td>-0.39024***</td>
<td>0.12142</td>
<td>0.677</td>
<td>-0.33346***</td>
<td>0.11406</td>
<td>0.716</td>
<td>-0.51342***</td>
<td>0.13585</td>
<td>0.598</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telecommunications</td>
<td>-0.4322***</td>
<td>0.14577</td>
<td>0.648</td>
<td>-0.44994***</td>
<td>0.13998</td>
<td>0.638</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real Estate</td>
<td>-2.57076***</td>
<td>0.88672</td>
<td>0.076</td>
<td>-2.02254***</td>
<td>0.88115</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility</td>
<td>-0.34602***</td>
<td>0.10332</td>
<td>0.707</td>
<td>-0.38452***</td>
<td>0.10302</td>
<td>0.681</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest and building products/homebuilders</td>
<td>-0.82258***</td>
<td>0.15075</td>
<td>0.439</td>
<td>-0.84771***</td>
<td>0.1487</td>
<td>0.428</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care / chemicals</td>
<td>-0.91114***</td>
<td>0.15151</td>
<td>0.402</td>
<td>-0.91742***</td>
<td>0.14909</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transportation</td>
<td>-0.21071*</td>
<td>0.11833</td>
<td>0.81</td>
<td>-0.35511**</td>
<td>0.1064</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High technology/computers/office equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** P-value ≤ 1%,  ** 1% < P-value ≤ 5%,  * 5% < P-value ≤ 10% based on Wald chi-square tests

#### PANEL B: MODEL FIT STATISTICS, 1984-2000

<table>
<thead>
<tr>
<th>Model</th>
<th>-2LogL</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downgrade TVC base model</td>
<td>18789.006</td>
<td>23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Downgrade TVC extended model</td>
<td>18538.62</td>
<td>32</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Up-grade TVC base model</td>
<td>34.7446</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Up-grade TVC extended model</td>
<td>40.85</td>
<td>5</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

#### PANEL C: SUMMARY OF THE NUMBER OF EVENTS AND CENSORED STATES, 1984-2000

<table>
<thead>
<tr>
<th>Stratum (Sequence of events)</th>
<th>Total</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1882</td>
<td>774</td>
<td>41.13</td>
<td>1108</td>
<td>58.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1115</td>
<td>445</td>
<td>39.91</td>
<td>670</td>
<td>60.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>654</td>
<td>238</td>
<td>36.39</td>
<td>416</td>
<td>63.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>374</td>
<td>125</td>
<td>33.42</td>
<td>249</td>
<td>66.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>47</td>
<td>21.36</td>
<td>173</td>
<td>78.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>116</td>
<td>32</td>
<td>27.59</td>
<td>84</td>
<td>72.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>15</td>
<td>25</td>
<td>45</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>10</td>
<td>30.3</td>
<td>23</td>
<td>69.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>6</td>
<td>35.29</td>
<td>11</td>
<td>64.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>71.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>33.33</td>
<td>2</td>
<td>66.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4487</td>
<td>1698</td>
<td>37.84</td>
<td>2789</td>
<td>62.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TVC Down-grade model

<table>
<thead>
<tr>
<th>Stratum (Sequence of events)</th>
<th>Total</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1882</td>
<td>774</td>
<td>41.13</td>
<td>1108</td>
<td>58.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1115</td>
<td>445</td>
<td>39.91</td>
<td>670</td>
<td>60.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>654</td>
<td>238</td>
<td>36.39</td>
<td>416</td>
<td>63.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>374</td>
<td>125</td>
<td>33.42</td>
<td>249</td>
<td>66.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>47</td>
<td>21.36</td>
<td>173</td>
<td>78.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>116</td>
<td>32</td>
<td>27.59</td>
<td>84</td>
<td>72.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>15</td>
<td>25</td>
<td>45</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>10</td>
<td>30.3</td>
<td>23</td>
<td>69.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>6</td>
<td>35.29</td>
<td>11</td>
<td>64.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>71.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>33.33</td>
<td>2</td>
<td>66.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4487</td>
<td>1698</td>
<td>37.84</td>
<td>2789</td>
<td>62.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TVC up-grade model

<table>
<thead>
<tr>
<th>Stratum (Sequence of events)</th>
<th>Total</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
<th>Event</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1882</td>
<td>774</td>
<td>41.13</td>
<td>1108</td>
<td>58.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1115</td>
<td>445</td>
<td>39.91</td>
<td>670</td>
<td>60.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>654</td>
<td>238</td>
<td>36.39</td>
<td>416</td>
<td>63.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>374</td>
<td>125</td>
<td>33.42</td>
<td>249</td>
<td>66.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>220</td>
<td>47</td>
<td>21.36</td>
<td>173</td>
<td>78.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>116</td>
<td>32</td>
<td>27.59</td>
<td>84</td>
<td>72.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>15</td>
<td>25</td>
<td>45</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>33</td>
<td>10</td>
<td>30.3</td>
<td>23</td>
<td>69.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>6</td>
<td>35.29</td>
<td>11</td>
<td>64.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>71.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>33.33</td>
<td>2</td>
<td>66.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4487</td>
<td>1698</td>
<td>37.84</td>
<td>2789</td>
<td>62.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison of the log-likelihood statistics in Panel B- Table 6.2 shows that adding the time interaction terms significantly improves upon the base models for both downgrades and upgrades. The likelihood ratio test shows that this improvement is significant at better than the 1 percent level.

The following discussion summarises the main findings relating to the TVC extended models. With respect to the significant industry and macro-economic covariates, the results of the TVC extended model are consistent with the results of the TVC base model. Similarly, the effects of rating history generally correspond to the effects found in the time-fixed model and the TVC base model. Consequently, the focus of the discussion is on the interaction of rating history with time, and the effects of rating history and time on migration hazards.

6.2.1. The effects of rating history and time interactions

The results of the main effect variables in the TVC extended models reinforce the key findings of the time-fixed and the TVC base models. Most rating history variables significant in the base models, with the exception of downgrade volatility (rate prior down), also appear as main effects in the extended models, and they retain the same sign. A few new main effects become significant in the extended model. For downgrades, the incidence of a big down jump in ratings (Number big down) is added and this variable increases the risk of a further downgrade. Rating age (Age since first rated) now appears in the upgrade model, where it reduces the probability of an upgrade. However, there are some important differences between the TVC base and the TVC extended models due to the presence of significant interaction covariates.

A notable change is the very large increase in the main effect of a downgrade at lag one (dummy lag1 down) on the risk of future downgrades. Without accounting for
time, a downgrade at lag one raises the hazard of a further downgrade by 148.6 percent, and this effect is constant as time progresses. After accounting for time, it increases the downgrade hazard by 308.9 percent at the beginning of a rating state. However, as time passes in the current rating, the time interaction for the lag one downgrade \((\text{lag1 down time})\) kicks in and the risk drops by 77 percent for every year the current rating remains unchanged. A downgrade at lag one then increases downgrade probability by \((4.089 \times 0.77^{1} - 1)\), or 214.85 percent at the end of year one \((\text{time} = 1)\) and 142.44 percent at the end of year two \((\text{time} = 2)\).

Most of the significant interaction terms have significant main effects in the extended models. However, a few interaction variables \((\text{Rate change time and rate down time})\) for downgrades, and \(\text{rate change time}\) for upgrades) have no significant main effect in the TVC extended model. Downgrade volatility \((\text{Rate prior down})\), however, does appear as one of the main effects in the TVC base and static time-fixed models.

Where a rating history variable is significant as a main effect in the extended model, the time interaction term has the opposite sign to the main effect. Consequently, the impact of the main effect decays as the duration of the current rating extends. For instance, without accounting for time, being in the investment rating boundary \((\text{dummy inv boundary})\) makes a downgrade 27.9 percent less likely. Taking into account time, the dummy variable reduces the downgrade probability by 39 percent at the commencement of the rating state. As time passes, the time interaction \(\text{inv boundary time}\), increases the downgrade hazard by 18.2 percent for each year the rating stays in its current grade. As a result, at the end of year one and year two, being in the investment boundary only decreases the downgrade hazard by 27.9 percent and 14.8 percent respectively.
It was hypothesised that issuers with a history of frequent downgrades (rate prior down) would display a tendency to descend to lower rating grades and would be less likely to travel upward. This holds before accounting for time interactions. After accounting for time interactions, rate prior down is not significant in the extended downgrade model. However, its interaction with time, rate down time, has a positive and significant coefficient so this increases the risk of a downgrade. By construction, rate down time is a decreasing function of time. Thus, as time unfolds, there is a declining effect.

Of particular interest, rate prior change is not significant in any model. However, its interaction with time, rate change time, is significant in the extended models for both downgrades and upgrades. The coefficient is negative for downgrades and positive for upgrades, therefore, the risk of a downgrade is reduced and the probability of an upgrade is increased. However, by construction rate change time is a declining function of time, so its impact decays with the passage of time.

Overall, employing a dynamic estimation framework that accounts for the evolution of macro-economic conditions over rating durations and accounting for interactions of rating history with time does not much alter the significant rating history variables that appear as main effects. However, there is strong evidence that many of these main effects decay with time and in some cases the change in the magnitude of the effect is very large.

6.3. Conclusion

This chapter extended the work developed in chapter 5 and examined the effects of rating history and time on migration hazards in a dynamic estimation framework. Two stratified time-varying covariate (TVC) hazard models were developed for each down
state and up state case in the estimation sample. Both the TVC base and the TVC extended models include firm-specific rating history variables, time-fixed and time-varying macro-economic covariates, and industry sector control variables. The dynamic estimation framework in the TVC models accounts for not only the state of the economic cycle at the beginning of each rating state but also changes in the macro-economic environment over time. In addition the TVC extended model incorporates the interaction terms between rating history and time, which are constructed as time-varying covariates and updated at each event time.

The objective of this chapter was to examine the following issues: (i) whether rating history variables remain significant determinants of rating migration hazards in a dynamic estimation framework, and (ii) whether time matters and how the effects of rating history on migration hazards are altered after accounting for time.

The results of the TVC base and the TVC extended models are consistent with the results of the time-fixed model. The dynamic estimation framework does not alter the key conclusions regarding the impact of rating history on migration hazards. Rating history is more important than current rating in determining future rating changes, particular for downgrades. There is evidence that certain aspects of rating history tend to repeat. Rating dynamics vary between downgrades and upgrades, with downgrades being more affected by rating history.

There is also strong evidence of significant interaction between rating history and time. The nature of the time interactions is that they reduce the main effect variables. In other words, the effects of rating history decay as the rating stays longer in its current grade. The results discussed in this chapter are in striking contrast to the time-homogeneous Markov assumption that all the effects of rating history is captured in the start rating and that rating migrations are time homogeneous.
Chapter 7

The predictive accuracy of rating history in a dynamic estimation framework

7.1. Introduction

This chapter investigates the predictive accuracy of rating history when forecasting rating migrations in a dynamic estimation framework. The forecast performance is analysed in terms of the Brier score and its skill components (bias, slope, scatter) at various forecast horizons.

The objective is twofold. First, to evaluate the forecast performance of the stratified dynamic Cox’s hazard base model with time-varying covariates (TVC base model) relative to the stratified Cox’s proportional hazard model (static time-fixed model). The question is, whether the dynamic estimation framework, which accounts for the evolution of macro-economic conditions over time, provides better predictions of rating changes. This is discussed in section 7.2 below.

The second objective is to assess the predictive performance of forecasts generated by the stratified dynamic Cox’s hazard extended model with time-varying covariates (TVC extended model) relative to the forecasts generated by the TVC base model. The question is, whether accounting for the interactions between rating history and the time for which the current rating persists improves the predictive accuracy of forecasts. This is discussed in section 7.3.

As discussed in chapter 4, the forecast technique results in a changing sample composition through time. Consequently, variations in forecast performance through time may be attributable to varying performance of the model, or to changing sample composition, or to some combination of the two. Comparisons across models at a
point in time can, however, readily be made because the sample is the same across
models at each point in time. A discussion about the sources of forecast errors and the
directions for future research is provided at the end of sections 7.2 and 7.3. Section
7.4 draws conclusions based on the findings of the empirical studies considered in this
chapter.

7.2. Stratified TVC base model versus stratified time-fixed model

The difference between the two stratified Cox’s hazard models lies in the construction
of the macro-economic covariates. In the time-fixed model, the lagged values of seven
macro-economic covariates were measured at the beginning of each rating state. The
static hazard model incorporates time-fixed covariates, which were not updated as the
rating continued in its current state. The static time-fixed model therefore only
accounts for the macro-economic conditions prevailing at the start of each
observation. On the other hand, the TVC base model employs both time-fixed and
time-varying macro-economic covariates. The lagged values of six time-varying
macro-economic covariates, except the recession index (dummy recession), were
measured at quarterly intervals. The values used in the estimation process were
updated to the most recent quarterly value as each risk set was formed. The recession
index (dummy recession) was formed as a time-fixed covariate with a value measured
at the beginning of each rating state. The TVC base model, therefore, controls for not
only the macro-economic environment prevailing at the start of each rating state but
also the development of macro-economic factors over the rating durations.

The following sections examine the relative forecast performance of the TVC base
model and the time-fixed model for downgrades and upgrades.
7.2.1. Downgrade models

7.2.1.1. Comparative forecast performance

The Brier scores of forecasts formed by the TVC base downgrade model are summarised in Panel A (short-term horizon) and Panel B (longer-term forecast horizon) of Table 7-1 below. For ease of reference, the Brier scores of the forecasts generated by the time-fixed downgrade model (Table 5-2, Panel A and Panel B) are appended to Table 7-1. Figure 7-1 depicts the covariance graphs of survival forecasts generated at the short-term horizon (Panel A, forecast time $t=0.5, 0.75$ and 1 year) and longer-term horizon (Panel B, forecast time $t=1, 2, and 3$ years).

In the aggregate, there is not much difference between the time-fixed and TVC base models. Both models exhibit a tendency to place most of their short-term estimates in probability categories of above 80 percent. The overuse of the optimistic categories, which are substantially higher than the relative survival frequencies, results in a positive bias within one-year forecast horizons (Panel A and Panel B Figure 7-1). At the longer-term forecast horizon, both time-fixed and TVC base models use optimistic categories far less often, and employ middle categories more often. The marked propensity to place a majority of survival forecasts in categories lower than the relative survival frequency translates into a negative bias at year two and thereafter (Panel B Figure 7-1). However, relative to the time-fixed model, the TVC base model uses optimistic categories for short-term estimates to a lesser extent, and uses middle and lower categories for longer-term estimates to a greater extent. As a result, the TVC base model achieves a smaller (positive) bias within the one-year forecast horizon and a larger (negative) bias at the longer horizon. Compared to the time-fixed model (Table 5-2), the TVC base model (Table 7-1) exhibits a tendency to switch
### Table 7-1
BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC BASE DOWNGRADE MODEL

<table>
<thead>
<tr>
<th>Forecast horizon t (year)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Naïve Brier score</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Benchmark Brier score</td>
<td>0.3670</td>
<td>0.2929</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2448</td>
<td>0.2199</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1222</td>
<td>0.0730</td>
</tr>
<tr>
<td>Forecast model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of forecasts</td>
<td>1872</td>
<td>1348</td>
</tr>
<tr>
<td>Mean probability of outcome</td>
<td>0.5721</td>
<td>0.6736</td>
</tr>
<tr>
<td>Mean probability of forecast</td>
<td>0.9327</td>
<td>0.8988</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.3641</td>
<td>0.2633</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2448</td>
<td>0.2199</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1193</td>
<td>0.0434</td>
</tr>
<tr>
<td>Forecast variance (scatter)</td>
<td>0.0041</td>
<td>0.0059</td>
</tr>
<tr>
<td>Reliability-in-the-large (bias square)</td>
<td>0.13</td>
<td>0.0507</td>
</tr>
<tr>
<td>2*Forecast-Outcome-Covariance</td>
<td>0.0148</td>
<td>0.0133</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0302</td>
<td>0.0302</td>
</tr>
</tbody>
</table>

### Table 5-2
BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TIME-FIXED DOWNGRADE MODEL

<table>
<thead>
<tr>
<th>Forecast horizon t (year)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Naïve Brier score</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Benchmark Brier score</td>
<td>0.3670</td>
<td>0.2929</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2448</td>
<td>0.2199</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1222</td>
<td>0.0730</td>
</tr>
<tr>
<td>Forecast model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of forecasts</td>
<td>1872</td>
<td>1348</td>
</tr>
<tr>
<td>Mean probability of outcome</td>
<td>0.5721</td>
<td>0.6736</td>
</tr>
<tr>
<td>Mean probability of forecast</td>
<td>0.9414</td>
<td>0.9163</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.3729</td>
<td>0.275</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2448</td>
<td>0.2199</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1281</td>
<td>0.0551</td>
</tr>
<tr>
<td>Forecast variance (scatter)</td>
<td>0.003</td>
<td>0.0048</td>
</tr>
<tr>
<td>Reliability-in-the-large (bias square)</td>
<td>0.1364</td>
<td>0.0589</td>
</tr>
<tr>
<td>2*Forecast-Outcome-Covariance</td>
<td>0.0114</td>
<td>0.0086</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0233</td>
<td>0.0196</td>
</tr>
</tbody>
</table>
Figure 7-1

BRIER SCORE COVARIANCE DECOMPOSITIONS OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC BASE DOWNGRADE MODEL

Panel A: Short-term forecasts

Brier score 0.2633, Scatter 0.0059, Bias 0.2252, Slope 0.0301

Panel B: Longer-term forecasts

Brier score 0.2741, Scatter 0.0279, Bias 0.2287, Slope 0.1038

Brier score 0.1772, Scatter 0.0326, Bias -0.1136, Slope 0.0611

Brier score 0.2403, Scatter 0.0411, Bias -0.3678, Slope 0.0689
more drastically from optimism in short-term survival estimates to pessimism in longer-term survival estimates.

Both the time-fixed and the TVC base models exhibit little variability (scatter) in estimates and seldom offer extremely pessimistic forecasts. Relative to the time-fixed model, the TVC base model exhibits more variability, due to the wider use of probability categories. On the index of slope, the prediction ability of both models is significantly dampened by the fact that a majority of short-term estimates for both survivors and non-survivors were placed into a few optimistic categories. However, at one- to three-year forecast horizons, the TVC base model is more discriminating and achieves a steeper slope.

Overall, the TVC base model performs slightly better within one-year forecast horizons, owing to a slight improvement in bias and slope. However, the relative performance of the TVC base model degenerates dramatically as the forecast horizon extends beyond year two and states of high downgrade risk leave the holdout sample.

7.2.1.2. Sources of forecast errors and implications

In the aggregate, the TVC base model does not perform particularly well compared to the time-fixed model. As suggested by Figlewski et al., (2008), the information contained in time-varying macro-economic variables is “incremental” compared to that contained in rating history alone. It was found that introducing time-varying macro-economic variables did not alter the effects of rating history variables on rating migrations. It appears that these variables did not adequately capture the effect of changing conditions out of sample on rating migrations. This can partly be attributed to the employment of static macro-economic data for holdout observations, as previously explained in Figure 4-4.
As suggested by Amato and Furfine (2003), the date of rating change is close to the time the actual credit review takes place. Thus, any decision by credit rating agencies is influenced by economic conditions at the time of the rating change. Furthermore, credit rating agencies have tended to adjust ratings downward more quickly in recent years (Altman and Kao, 1991), and show excessive pessimism in economic downturns (Amato and Furfine, 2003). Down states are typically short-lived and a majority of downgrades mass at survival durations varying from a quarter to one year (see Figure 4-15). As down states are subject to short credit review cycles, it is not surprising that the updating of macro-economic covariates gives the TVC base model a small information advantage at the short horizons.

Relative to the time-fixed model (Table 5-1 Panel A), the TVC base model (Table 6-1 Panel A) features a greater number of macro-economic covariates. In particular, it contains variables that capture being in a recession (dummy recession), industrial production change, and stock market performance (SP500 quarterly return). The presence of these time-varying macro-economic variables results in higher downgrade probability estimates and, accordingly, lower survival estimates, given the economic downturn in the holdout period, 2001-2005. Consequently, the TVC base model generates survival estimates that are less optimistic in the short-term and more pessimistic in the longer-term than those formed by the static time-fixed model. The TVC base model, therefore, has a smaller (positive) bias in short-term survival estimates and a larger (negative) bias in longer-term survival estimates.
7.2.2. Upgrade models

7.2.2.1. Comparative forecast performance

The Brier scores of forecasts formed by the TVC base upgrade model are summarised in Panel A (short-term horizon) and Panel B (longer-term horizon) of Table 7-2 below. For ease of reference, the Brier scores of the forecasts generated by the static time fixed upgrade model (Table 5-3, Panel A and Panel B) are appended to Table 7-2. Figure 7-2 depicts the covariance graphs of survival forecasts generated at the short-term horizon (Panel A, forecast time \( t=0.5, 0.75 \) and 1 year) and longer-term horizon (Panel B, forecast time \( t=1, 2, \) and 3 years).

In aggregate, there is not much difference in the short-term forecast performance of the two upgrade models. Relative to the time-fixed model, the TVC base model achieves similar skill components within the one-year forecast horizons. Both upgrade models have minimal variability and small bias in short-term survival estimates. However, both models perform poorly in discriminating rating states that survived from states that are upgraded (non-survivors).

Reference to the covariance graphs in Panel A and Panel B Figure 7-2 shows that the preponderance of optimistic estimates was most evident within the one-year forecast horizon. Like the time-fixed upgrade model, the TVC model places no estimate in the pessimistic categories below 40 percent. Both the time-fixed and the TVC models display a propensity to assign short-term survivor forecasts to the categories slightly above the relative survivor frequency. This translates into a small positive bias, particularly at the one- and two-year forecast horizon.
Table 7-2
BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC BASE UPGRADE MODEL

<table>
<thead>
<tr>
<th>Forecast horizon t (year)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>Naïve Brier score</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Benchmark Brier score</td>
<td>0.129</td>
<td>0.1493</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.1137</td>
<td>0.13</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0153</td>
<td>0.0193</td>
</tr>
</tbody>
</table>

Forecast model

<table>
<thead>
<tr>
<th>Forecast variance (scatter)</th>
<th>0.0153</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast variance (scatter)</td>
<td>0.0009</td>
</tr>
<tr>
<td>Reliability-in-the-large (bias square)</td>
<td>0.0091</td>
</tr>
<tr>
<td>2^Forecast-Outcome-Covariance</td>
<td>0.0018</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0079</td>
</tr>
</tbody>
</table>

Table 5-3
BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TIME-FIXED UPGRADE MODEL

<table>
<thead>
<tr>
<th>Forecast horizon t (year)</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>Naïve Brier score</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Benchmark Brier score</td>
<td>0.129</td>
<td>0.1493</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.1137</td>
<td>0.13</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0153</td>
<td>0.0193</td>
</tr>
</tbody>
</table>

Forecast model

<table>
<thead>
<tr>
<th>Forecast variance (scatter)</th>
<th>0.0153</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast variance (scatter)</td>
<td>0.0007</td>
</tr>
<tr>
<td>Reliability-in-the-large (bias square)</td>
<td>0.0008</td>
</tr>
<tr>
<td>2^Forecast-Outcome-Covariance</td>
<td>0.0015</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0066</td>
</tr>
</tbody>
</table>

165
Figure 7-2

BRIER SCORE COVARIANCE DECOMPOSITIONS OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC BASE UPGRADE MODEL

Panel A: Short-term forecasts

- Brier score 0.1393, Scatter 0.0012, Bias 0.098, Slope 0.0062
- Brier score 0.1528, Scatter 0.0018, Bias 0.1015, Slope 0.0125
- Brier score 0.1579, Scatter 0.003, Bias 0.0917, Slope 0.026

Panel B: Longer-term forecasts

- Brier score 0.1126, Scatter 0.0023, Bias 0.0361, Slope 0.0214
- Brier score 0.1528, Scatter 0.0018, Bias 0.1015, Slope 0.0125
- Brier score 0.1579, Scatter 0.003, Bias 0.0917, Slope 0.026
The forecast performances of the TVC base model and the time-fixed model deteriorate at the longer-term horizons. Both models underestimate longer-term survival forecasts, resulting in negative bias. The TVC base model uses middle and lower categories to a lesser extent, and employs optimistic categories to a greater extent than did the time-fixed model. It switches from optimism in short-term survival estimates to pessimism in longer-term survival estimates in a less drastic fashion than does the time-fixed model.

7.2.2.2. Sources of forecast errors and implications

The TVC upgrade base model, compared to the time-fixed upgrade model, features similar rating history covariates and fewer macro-economic covariates. A notable change is that the output growth gap ($\text{RealGDP}_\text{actual} - \text{potential}$) has an opposite effect in the two models. A decrease in the output growth gap makes an upgrade more likely in the time-fixed model (Table 5-1 Panel A) but less likely in the TVC base model (Table 6-1 Panel A). Given the unfavourable performance of the output growth gap during the economic downturn in 2001-2005, survival (no-upgrade) estimates generated by the TVC base model are higher than survival (no-upgrade) estimates formed by the time-fixed model. The TVC base model, accordingly, has a more (positive) bias in short-term estimates and a less (negative) bias in longer-term estimates relative to the time-fixed model. This is in contrast to the comparative forecast performance of the downgrade models discussed in section 7.1.

The edge of the TVC upgrade base model in forecast performance at longer-term horizons can be attributed to the model’s better calibrated forecasts: the survival estimates locate closer to the relative survival frequency.
7.3. Stratified TVC extended model versus stratified TVC base model

This section examines the comparative forecast performance of the survival estimates generated by the TVC extended model and the TVC base model. Both stratified TVC extended and TVC base models control for the evolution of macro-economic conditions over survival durations. The difference lies in the presence of time-varying covariates that capture the interactions between rating history and survival time in the TVC extended model.

As discussed in chapter 4, when the time interaction terms are introduced into the models only conditional forecasts can be made. The \( t \)-period survival forecasts generated by the TVC extended model are constructed conditional on holdout states surviving at the time horizon \( T^* \), which can take on different values over the forecast horizon \( t \). The time interaction terms of holdout state \( q \), \( Z^q_p(t) \), which are used in Equations (9) to form \( t \)-period estimates, are constructed as in Equations (16), (17), (18), and (19) at the conditional survival time \( T^* \).

The presence of the interaction terms between rating history and time introduces two effects on the conditional survival estimates: the decay effect and the information advantage effect, as discussed in chapter 4. The decay effect is driven by the extension of time, and the impact on the survival probability is affected by both the magnitude and sign of the main effects. As the rating durations get longer, rating history variables interact with time, and their main effects decay with time. On the other hand, the information advantage effect is driven by the changing sample composition. As time extends and issuers change their rating grade, they are deleted from the holdout sample. Consequently, the predictive accuracy of \( t \)-period
conditional survival estimates is influenced by both the decay effect and the information advantage effect.

The following discussion examines the comparative forecast performance of the TVC extended model and the TVC base model for downgrades and upgrades. An analysis of the decay effect and the information advantage effect provides insight into the variations of survival forecasts generated by the two dynamic TVC models.

7.3.1. Downgrade models

7.3.1.1. Comparative forecast performance

Table 7-3 below summarises the Brier scores and skill components of survival forecasts generated by the time-fixed, TVC base, and TVC extended downgrade models at various forecast horizons.

The covariance decompositions of the Brier scores of conditional survival forecasts formed by the TVC extended downgrade model are summarised in Panel A (short-term horizon) and Panel B –E (longer-term horizon) of Table 7-4 below.

Figure 7-3 depicts the covariance graphs of survival forecasts conditional on holdout states surviving at time $T^*=0.25\ \text{year}$ (Panel A, forecast horizon $t=0.5, 0.75$ and 1 year), at time $T^*=1\ \text{year}$ (Panel B, forecast horizon $t= 2, \text{and } 3\ \text{years}$), and at time $T^*=2\ \text{years}$ (Panel C, forecast horizon $t= 3\ \text{years}$).

The following discussion analyses the performance of conditional survival forecasts. Focus is given to the conditional surviving time $T^*=0.25\ \text{year}$ and $T^*=1\ \text{year}$, as these times capture the short durations of down states. Most downgrades are observed within a one-year window and the one-year forecast horizon is of natural interest to financial institutions and regulators.
<table>
<thead>
<tr>
<th>Forecast horizon $t$ (year)</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Brier score*</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Benchmark Brier score**</td>
<td>0.3670</td>
<td>0.2929</td>
<td>0.2447</td>
<td>0.2038</td>
<td>0.3012</td>
<td>0.1539</td>
<td>0.0769</td>
<td>0.0613</td>
<td>0.2015</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1222</td>
<td>0.0730</td>
<td>0.0498</td>
<td>0.0302</td>
<td>0.0564</td>
<td>0.0038</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0240</td>
</tr>
<tr>
<td>Stratified proportional model (time-fixed model)</td>
<td>0.3729</td>
<td>0.275</td>
<td>0.2203</td>
<td>0.1908</td>
<td>0.3001</td>
<td>0.1714</td>
<td>0.2109</td>
<td>0.3919</td>
<td>0.5816</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1281</td>
<td>0.0551</td>
<td>0.0254</td>
<td>0.0172</td>
<td>0.0553</td>
<td>0.0213</td>
<td>0.1368</td>
<td>0.3333</td>
<td>0.4041</td>
</tr>
<tr>
<td>Stratified TVC base model</td>
<td>0.3641</td>
<td>0.2633</td>
<td>0.21</td>
<td>0.1834</td>
<td>0.2741</td>
<td>0.1772</td>
<td>0.2403</td>
<td>0.4714</td>
<td>0.6828</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1193</td>
<td>0.0434</td>
<td>0.0151</td>
<td>0.0098</td>
<td>0.0293</td>
<td>0.0271</td>
<td>0.1662</td>
<td>0.4128</td>
<td>0.5053</td>
</tr>
<tr>
<td>Stratified TVC extended model</td>
<td>Forecasts conditional on firms surviving at $T^* = 0.25$ year</td>
<td>0.2882</td>
<td>0.2268</td>
<td>0.1904</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0683</td>
<td>0.0319</td>
<td>0.0168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^* = 1$ year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^* = 2$ years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^* = 3$ years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^* = 4$ years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The naïve Brier score is constructed by setting the probability forecast at time $t$ in equation (11) equal to a random forecast of 0.5.

** The Brier score of the benchmark was constructed by setting the probability forecast at time $t$ in equation (11) equal to the proportion of rating states surviving beyond time $t$ in the estimation sample.
**Table 7-4**

BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC EXTENDED DOWNGRADE MODEL

<table>
<thead>
<tr>
<th>Forecast horizon t (year)</th>
<th>Panel A: Survival forecasts conditional on firms surviving at time T (^*) = 0.25 year</th>
<th>Panel B: Survival forecasts conditional on firms surviving at T (^*) = 1 year</th>
<th>Panel C: Survival forecasts conditional on firms surviving at T (^*) = 2 years</th>
<th>Panel D: Survival forecasts conditional on firms surviving at T (^*) = 3 years</th>
<th>Panel E: Survival forecasts conditional on firms surviving at T (^*) = 4 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Brier score</td>
<td>0.25 0.25 0.25</td>
<td>0.25 0.25 0.25</td>
<td>0.25 0.25 0.25</td>
<td>0.25 0.25 0.25</td>
<td>0.25 0.25 0.25</td>
</tr>
<tr>
<td>Benchmark Brier score</td>
<td>0.2929 0.2447 0.2038</td>
<td>0.1539 0.0769 0.0613</td>
<td>0.2015</td>
<td>0.0696 0.0741 0.0586</td>
<td>0.1775</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.2199 0.1949 0.1736</td>
<td>0.1501 0.0741 0.0586</td>
<td>0.1775</td>
<td>0.0741 0.0586 0.0401</td>
<td>0.1775</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0730 0.0498 0.0302</td>
<td>0.0038 0.0028 0.0027</td>
<td>0.0240</td>
<td>0.0028 0.0027 0.0027</td>
<td>0.0240</td>
</tr>
</tbody>
</table>

**Forecast model**

| Number of forecasts      | 1348 1059 832                                 | 669 273 80 13                                  | 273 80 13                                     | 80 13                                           | 13                                            |
| Mean probability of outcome | 0.6736 0.7347 0.7764                           | 0.6161 0.9194 0.9375 0.7692                   | 0.9194 0.0311 0.0128                         | 0.0311 0.0436 0.0001                           | 0.0001                                        |
| Mean probability of forecast | 0.937 0.911 0.8995                            | 0.7981 0.7323 0.6691 0.7805                   | 0.0033 0.035 0.072 0.0001                     | 0.0035 0.035 0.072 0.0001                       | 0.0001                                        |
| Brier score              | 0.2882 0.2268 0.1904                           | 0.1699 0.1534 0.1841 0.2305                   | 0.0003 0.035 0.072 0.0001                     | 0.0003 0.035 0.072 0.0001                       | 0.0001                                        |
| Outcome index variance   | 0.2199 0.1949 0.1736                            | 0.1501 0.0741 0.0586                            | 0.1775                                        | 0.0741 0.0586 0.0401                            | 0.1775                                        |
| Skill components         | 0.0683 0.0319 0.0168                            | 0.0198 0.0793 0.1255 0.053                     | 0.0028 0.0035 0.0014                          | 0.0035 0.035 0.072 0.0001                       | 0.0001                                        |
| Forecast variance (scatter) | 0.0037 0.007 0.0102                           | 0.0279 0.0391 0.0463 0.0088                   | 0.0003 0.035 0.072 0.0001                     | 0.0035 0.035 0.072 0.0001                       | 0.0001                                        |
| Reliability-in-the-large (bias square) | 0.0694 0.0311 0.0128 | 0.0033 0.035 0.072 0.0001 | 0.0003 0.035 0.072 0.0001 | 0.0035 0.035 0.072 0.0001 | 0.0001 |
| 2*Forecast-Outcome-Covariance | 0.0048 0.0062 0.0061 | 0.0008 -0.0002 -0.0002 | 0.0008 -0.0002 -0.0002 | 0.0008 -0.0002 -0.0002 | 0.0008 -0.0002 -0.0002 |
| Slope                    | 0.0109 0.0159 0.0176                            | 0.0280 -0.0351 -0.0614 -0.1239                 | 0.0280 -0.0351 -0.0614 -0.1239                | 0.0280 -0.0351 -0.0614 -0.1239                 | 0.0280 -0.0351 -0.0614 -0.1239                |

Panel A: Survival forecasts conditional on firms surviving at time T \(^*\) = 0.25 year
Panel B: Survival forecasts conditional on firms surviving at T \(^*\) = 1 year
Panel C: Survival forecasts conditional on firms surviving at T \(^*\) = 2 years
Panel D: Survival forecasts conditional on firms surviving at T \(^*\) = 3 years
Panel E: Survival forecasts conditional on firms surviving at T \(^*\) = 4 years
Figure 7-3

BRIER SCORE COVARIANCE DECOMPOSITIONS OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC EXTENDED DOWNGRADE MODEL

Panel A: Short-term forecasts conditional on firms surviving at $T^*=0.25$ year

Forecasts at 0.5 year lead time conditional on firm surviving at 0.25 year lead time

- Mean survival outcome index
- Estimated probability of survival
- Migration frequency histogram
- Survival frequency histogram
- Mean survival probabilistic forecast
- Regression line for forecasts on outcome indexes

Brier score 0.2882, Scatter 0.0037, Bias 0.2634, Slope 0.0109

Forecasts at 0.75 year lead time conditional on firm surviving at 0.25 year lead time

- Mean survival outcome index
- Estimated probability of survival
- Migration frequency histogram
- Survival frequency histogram
- Mean survival probabilistic forecast
- Regression line for forecasts on outcome indexes

Brier score 0.2268, Scatter 0.007, Bias 0.1764, Slope 0.0162

Forecasts at 1 year lead time conditional on firm surviving at 0.25 year lead time

- Mean survival outcome index
- Estimated probability of survival
- Migration frequency histogram
- Survival frequency histogram
- Mean survival probabilistic forecast
- Regression line for forecasts on outcome indexes

Brier score 0.1904, Scatter 0.0102, Bias 0.1131, Slope 0.0175

Panel B: Longer-term forecasts conditional on firms surviving at $T^*=1$ year

Forecasts at 2 year lead time conditional on firm surviving at 1 year lead time

- Mean survival outcome index
- Estimated probability of survival
- Migration frequency histogram
- Survival frequency histogram
- Mean survival probabilistic forecast
- Regression line for forecasts on outcome indexes

Brier score 0.1699, Scatter 0.0279, Bias -0.0173, Slope 0.0279

Forecasts at 3 year lead time conditional on firm surviving at 1 year lead time

- Mean survival outcome index
- Estimated probability of survival
- Migration frequency histogram
- Survival frequency histogram
- Mean survival probabilistic forecast
- Regression line for forecasts on outcome indexes

Brier score 0.1534, Scatter 0.0391, Bias -0.187, Slope -0.035

Panel C: Longer-term forecasts conditional on firms surviving at $T^*=2$ years

Forecasts at 3 year lead time conditional on firm surviving at 2 year lead time

- Mean survival outcome index
- Estimated probability of survival
- Migration frequency histogram
- Survival frequency histogram
- Mean survival probabilistic forecast
- Regression line for forecasts on outcome indexes

Brier score 0.19, Scatter 0.0522, Bias -0.2369, Slope -0.052
In general, the TVC extended model uses optimistic probability categories more often, and employs pessimistic categories less often, than the TVC base model. The conditional survival forecasts are therefore higher, i.e. more optimistic in the short term and less pessimistic in the longer term, than the corresponding unconditional forecasts formed by the TVC base model. However, both TVC downgrade models show poor discrimination.

As shown in Table 7-3, the short-term survival estimates conditional on surviving time $T^* = 0.25$ year do not perform as well as the corresponding unconditional survival estimates. This underperformance is due to deterioration in bias and a flat slope. Both TVC downgrade models exhibit a strong tendency to form estimates that are substantially higher than the relative survival frequencies. However, the TVC extended model (Panel A Figure 7-3) overestimates short-term survival forecasts to a greater extent than the TVC base model (Panel A Figure 7-1). This results in a larger (positive) bias. Both models achieve negligible scatter as a majority of short-term estimates were placed in a single probability category (above 90 percent).

As shown in Table 7-3, the longer-term estimates generated by the TVC extended model conditional on surviving time $T^* = 1$ year and $T^* = 2$ years perform better than the respective unconditional estimates generated by the TVC base model. This improvement in forecast accuracy can be attributed to a smaller negative bias in conditional survival estimates (Panel B and C Figure 7-3). The striking feature in Panel B Figure 7-3 is that the TVC extended model offers well-calibrated survival estimates at the two-year horizon, as evidenced by a minimal bias. This is a substantial improvement compared to the performance of the TVC base model at the same horizon (Panel B Figure 7-1).
7.3.1.2. Sources of differences in forecast performance

The following discussion analyses the decay effect and the information advantage effect on survival forecasts conditional on holdout states surviving at $T^*=0.25$ year and $T^*=1$ year.

**a. Decay effect**

The decay effect reduces the main effects of rating history variables on migration hazard as the current rating duration gets longer. For the purpose of illustration, the decay effects of two variables, a downgrade at lag one rating state (dummy lag1 down) and an unrated period in rating history (Number NR), on the hazard of a downgrade are depicted below. These two rating history variables have opposite effects on the hazard of a downgrade and so do their decay effects.

**Figure 7-4**

As discussed in chapter 6 (Panel A Table 6-2) and as depicted in Figure 7-4 above, without accounting for the time-interaction (TVC base model) a downgrade at lag one (dummy lag1 down) increases the risk of a subsequent downgrade by 149 percent.
After accounting for time-interaction (TVC extended model), the risk increases by 309 percent at the start of the new rating, and that risk then decays so that by the end of the first quarter, the first year, and the second year, it has fallen to 283 percent, 215 percent and 142 percent respectively.

The decay effects of dummy lag1 down and other variables that raise the risk of a downgrade translate into more optimistic survival estimates as the forecast horizon extends. Accounting for the decay effects of these variables, the conditional survival estimates, compared to the unconditional estimates, should be lower (more pessimistic) at the short end (time zero) and higher (more optimistic) at the longer end.

**Figure 7-5**

<table>
<thead>
<tr>
<th>Changes in downgrade risk</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>-54.8%</td>
<td>0</td>
</tr>
<tr>
<td>-51.9%</td>
<td>0.25</td>
</tr>
<tr>
<td>-42.2%</td>
<td>0.5</td>
</tr>
<tr>
<td>-40.5%</td>
<td>0.75</td>
</tr>
<tr>
<td>-40.5%</td>
<td>1</td>
</tr>
<tr>
<td>-26.2%</td>
<td>1.25</td>
</tr>
<tr>
<td>-40.5%</td>
<td>1.5</td>
</tr>
<tr>
<td>-40.5%</td>
<td>1.75</td>
</tr>
<tr>
<td>-40.5%</td>
<td>2</td>
</tr>
</tbody>
</table>

On the other hand, as depicted in Figure 7-5 above, the occurrence of an unrated period (Number NR) makes a downgrade 40.5 percent less likely without accounting for the time interaction (TVC base model). After taking into account the interaction between Number NR and time (TVC extended model), the hazard of a downgrade is
diminished by 54.8 percent at the beginning of a rating state. As time progresses, the
hazard of a downgrade is reduced by 51.9 percent, 42.2 percent, and 26.2 percent at
one-quarter, one-year and two-year lead time respectively.

The decay effects of number NR and other variables that diminish the hazard of a
downgrade result in more pessimistic survival estimates as time progresses.
Accounting for the decay effects of these variables, the conditional survival estimates,
relative to the unconditional estimates, should be higher (more optimistic) in the
short-term (time zero) and lower (more pessimistic) in the longer-term horizons.

However, several factors cancel out and blur the decay effects of individual rating
history variables such as dummy lag1 down and number NR. First, the magnitude of a
variable can amplify the magnitude of its hazard ratio substantially. For instance, the
current rating (start rating) can take on values varying from 1 to 26, with the median
equal to 12 and the mean equal to 12.73. Lag one and lag two take on values up to
4.69 and 4.72 years respectively, whereas dummy lag1 down can takes on values of
either 1 or 0. The decay effect is, therefore, influenced by not only the sign and size of
the coefficient but also the magnitude of the main effects. Second, for quarterly
forecasts conditional on survival time $T^*=0.25$ year, the forecast horizon extends at
quarterly intervals, and these shorter time periods result in smaller decay effects on
survival estimates. However, for yearly forecasts conditional on survival time $T^*=1$
year, the forecast horizon extends at yearly intervals and the effects of decay on
survival estimates become more pronounced.

As shown in Panel A Table 7-4 and Table 7-1, conditional forecasts are more
optimistic and have larger positive bias at the short-term horizon. This might be due
to the decay effects being dominated by the interaction between time and variables
with a negative sign (reducing migration risk) and a relatively large magnitude, such
as the current rating (start rating). However, it is difficult to disentangle the decay effects from the effects of changing sample composition.

**b. Information advantage effect**

The conditional survival time $T^*$ embedded in the interaction terms could convey useful information about the expected life of holdout states, and bring about the information advantage effect. This depends on how informative the conditional survival time $T^*$ is. For instance, the conditional time $T^*$=0.25 years conveys little information that is useful for allocating surviving states and downgraded states to the right probability categories. Knowing that a rating grade survived at a one-quarter lead time ($T^*$=0.25 year) provides little information about whether the rating will descend or ascend the rating spectrum, or retain the current rating grade. On the other hand, the conditional time $T^*$=1 year conveys useful information, as states surviving at one-year lead time are more likely to stay in the current rating grade or go up the rating grades over the forecast horizon. It is, therefore, not surprising that the two-year forecasts conditional on issuers surviving at $T^*$=1 year have minimal bias.

As shown in Figure 7-3 Panel B and Figure 7-1 Panel B, longer-term forecasts conditional on holdout states surviving at $T^*$=1 year achieve smaller negative bias compared to the corresponding unconditional survival forecasts. This may be attributed to the information advantage embedded in the interaction terms between rating history and the conditional survival time $T^*$.

**7.3.2. Upgrade models**

**7.3.2.1. Comparative forecast performance**

Table 7-5 summarises the Brier scores and skill components of survival forecasts generated by the time-fixed model, TVC base model and TVC extended upgrade
model at various forecast horizons.

The Brier scores of conditional survival forecasts formed by the TVC extended upgrade model are summarised in Table 7-6 Panel A (short-term horizon) and Panel B –Panel E (longer-term horizon).

Figure 7-6 depicts the covariance graphs of survival forecasts conditional on holdout states surviving at time $T^*=0.25$ year (Panel A, forecast horizon $t=0.5, 0.75$ and 1 year), at time $T^*=1$ year (Panel B, forecast horizon $t= 2$, and 3 years), and at time $T^*=2$ years (Panel C, forecast horizon $t= 3$ years).

The following discussion analyses the performance of conditional survival forecasts. Focus is given to the conditional survival time $T^*=0.25$ year, $T^*=1$ year and $T^*=2$ years, as most upgrades are observed within a two-year window.

The TVC extended upgrade model, like the time-fixed and TVC base upgrade models, achieves minimal bias and offers well-calibrated survival estimates within the two-year forecast horizon. In general, the TVC extended model uses optimistic probability categories to a lesser extent and employs the middle probability categories to a greater extent than does the TVC base model. The TVC extended model, therefore, achieves a smaller positive bias in short-term survival estimates and a larger negative bias in longer-term survival estimates.
Table 7-5

BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY STRATIFIED UPGRADE MODELS

<table>
<thead>
<tr>
<th>Forecast horizon t (year)</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naïve Brier score</strong>#</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Benchmark Brier score</strong>##</td>
<td>0.129</td>
<td>0.1493</td>
<td>0.1671</td>
<td>0.1772</td>
<td>0.1178</td>
<td>0.1749</td>
<td>0.1225</td>
<td>0.0698</td>
<td>0.0736</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0153</td>
<td>0.0193</td>
<td>0.0229</td>
<td>0.0226</td>
<td>0.0041</td>
<td>0.0068</td>
<td>0.0001</td>
<td>0.0004</td>
<td>0.0026</td>
</tr>
<tr>
<td><strong>Stratified proportional model (time-fixed model)</strong></td>
<td>0.1221</td>
<td>0.1393</td>
<td>0.1527</td>
<td>0.1582</td>
<td>0.1124</td>
<td>0.1593</td>
<td>0.1704</td>
<td>0.243</td>
<td>0.3602</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0084</td>
<td>0.0093</td>
<td>0.0085</td>
<td>0.0036</td>
<td>-0.0013</td>
<td>-0.0088</td>
<td>0.048</td>
<td>0.1736</td>
<td>0.2892</td>
</tr>
<tr>
<td><strong>Stratified TVC base model</strong></td>
<td>0.1219</td>
<td>0.1393</td>
<td>0.1528</td>
<td>0.1579</td>
<td>0.1126</td>
<td>0.1631</td>
<td>0.1522</td>
<td>0.1453</td>
<td>0.1433</td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0082</td>
<td>0.0093</td>
<td>0.0086</td>
<td>0.0033</td>
<td>-0.0011</td>
<td>-0.0050</td>
<td>0.0298</td>
<td>0.0759</td>
<td>0.0723</td>
</tr>
<tr>
<td><strong>Stratified TVC extended model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^*=$0.25 year</td>
<td>0.1278</td>
<td>0.1379</td>
<td>0.1471</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>-0.0022</td>
<td>-0.0063</td>
<td>-0.0075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^*=$1 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>-0.0045</td>
<td>0.049</td>
<td>0.1295</td>
<td>0.1865</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^*=$2 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1745</td>
<td>0.1949</td>
<td>0.2279</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^*=$3 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>0.0521</td>
<td>0.1255</td>
<td>0.1569</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^*=$4 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>0.2091</td>
<td>0.2372</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts conditional on firms surviving at $T^*=$4 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill components</td>
<td>0.1397</td>
<td>0.1662</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# The naïve Brier score is constructed by setting the probability forecast at time $t$ in equation (11) equal to a random forecast of 0.5.

## The Brier score of the benchmark was constructed by setting the probability forecast at time $t$ in equation (11) equal to the proportion of rating states surviving beyond time $t$ in the estimation sample.
### Table 7-6

**BRIER SCORE OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC EXTENDED UPGRADE MODEL**

<table>
<thead>
<tr>
<th>Forecast horizon $t$ (year)</th>
<th>Panel A: Survival forecasts conditional on firms surviving at time $T^*=0.25$ year</th>
<th>Panel B: Survival forecasts conditional on firms surviving at $T^*=1$ year</th>
<th>Panel C: Survival forecasts conditional on firms surviving at $T^*=2$ years</th>
<th>Panel D: Survival forecasts conditional on firms surviving at $T^*=3$ years</th>
<th>Panel E: Survival forecasts conditional on firms surviving at $T^*=4$ years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>0.75</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Naïve Brier score</strong></td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Benchmark Brier score</strong></td>
<td>0.1493</td>
<td>0.1671</td>
<td>0.1772</td>
<td>0.1749</td>
<td>0.1225</td>
</tr>
<tr>
<td><strong>Outcome index variance</strong></td>
<td>0.13</td>
<td>0.1442</td>
<td>0.1546</td>
<td>0.1681</td>
<td>0.1224</td>
</tr>
<tr>
<td><strong>Skill components</strong></td>
<td>0.0193</td>
<td>0.0229</td>
<td>0.0226</td>
<td>0.0068</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Forecast model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of forecasts</td>
<td>1348</td>
<td>1059</td>
<td>832</td>
<td>669</td>
<td>273</td>
</tr>
<tr>
<td>Mean probability of outcome</td>
<td>0.8464</td>
<td>0.8253</td>
<td>0.8089</td>
<td>0.7862</td>
<td>0.8571</td>
</tr>
<tr>
<td>Mean probability of forecast</td>
<td>0.9048</td>
<td>0.8728</td>
<td>0.8459</td>
<td>0.7513</td>
<td>0.6752</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.1278</td>
<td>0.1379</td>
<td>0.1471</td>
<td>0.1636</td>
<td>0.1714</td>
</tr>
<tr>
<td>Outcome index variance</td>
<td>0.13</td>
<td>0.1442</td>
<td>0.1546</td>
<td>0.1681</td>
<td>0.1224</td>
</tr>
<tr>
<td><strong>Skill components</strong></td>
<td>-0.0022</td>
<td>-0.0063</td>
<td>-0.0075</td>
<td>-0.0045</td>
<td>0.049</td>
</tr>
<tr>
<td>Forecast variance (scatter)</td>
<td>0.0052</td>
<td>0.0092</td>
<td>0.0126</td>
<td>0.019</td>
<td>0.027</td>
</tr>
<tr>
<td>Reliability-in-the-large (bias square)</td>
<td>0.0034</td>
<td>0.0023</td>
<td>0.0014</td>
<td>0.0012</td>
<td>0.0331</td>
</tr>
<tr>
<td>2*Forecast-Outcome-Covariance</td>
<td>0.0107</td>
<td>0.0178</td>
<td>0.0214</td>
<td>0.0247</td>
<td>0.0111</td>
</tr>
<tr>
<td>Slope</td>
<td>0.0412</td>
<td>0.0617</td>
<td>0.0692</td>
<td>0.0735</td>
<td>0.0453</td>
</tr>
</tbody>
</table>

Panel A: Survival forecasts conditional on firms surviving at time $T^*=0.25$ year
Panel B: Survival forecasts conditional on firms surviving at $T^*=1$ year
Panel C: Survival forecasts conditional on firms surviving at $T^*=2$ years
Panel D: Survival forecasts conditional on firms surviving at $T^*=3$ years
Panel E: Survival forecasts conditional on firms surviving at $T^*=4$ years

Panel F: Survival forecasts conditional on firms surviving at $T^*=5$ years
Figure 7-6

BRIER SCORE COVARIANCE DECOMPOSITIONS OF SURVIVAL FORECASTS GENERATED BY THE STRATIFIED TVC EXTENDED UPGRADE MODEL

Panel A: Short-term forecasts conditional on firms surviving at T*=0.25 year

Panel B: Longer-term forecasts conditional on firms surviving at T*=1 year

Panel C: Longer-term forecasts conditional on firms surviving at T*=2 years
The short-term survival forecasts conditional on surviving time $T^* = 0.25$ year (Table 7-5) perform better than the respective unconditional forecasts generated by the TVC base model. This improvement in forecast accuracy is due to a smaller positive bias and a shallow slope. Relative to the TVC base model, the TVC extended model exhibits an edge in matching mean survival estimates to relative survival frequency within the one-year forecast window. As depicted in Panel A Figure 7-6, the mean forecast line and the survival base rate line cross slightly above the 45-degree line. This corresponds to a minimal bias in short-term conditional forecasts.

At the longer-term forecast horizon, the conditional survival estimates (Table 7-5) perform poorly relative to the corresponding unconditional survival estimates. The inferior forecast performance of the TVC extended model is attributable to a larger negative bias and more variability in the estimates. Both TVC upgrade models show excessive pessimism in assigning longer-term survival estimates. However, the TVC extended model underestimates survival forecasts by a larger margin (Panel B and Panel C Figure 7-6). In addition, the extended model exhibits more variability in the estimates due to the broader use of probability categories.

7.3.2.2. Sources of differences in forecast performance

a. Decay effect

As shown in Table 7-6 and Table 7-2, conditional TVC forecasts are less optimistic, and fall below unconditional TVC forecasts across all horizons. Survival forecasts conditional on $T^* = 0.25$ year (Panel A Table 7-6) have smaller positive bias than the corresponding unconditional forecasts (Panel A Table 7-2). The decay effects appear to become more pronounced as the forecast horizon extends by yearly intervals ($T^* = 1$ year, $T^* = 2$ years). Consequently, the TVC extended model has a larger negative bias
in longer-term estimates than the TVC base model. However this may also be affected by changes in the sample composition.

b. Information advantage effect

As mentioned earlier, the information advantage effect is driven by the sample composition change. Unlike down states, up states are resilient and have longer survival durations. A substantial proportion of holdout states survive for one quarter and a majority of holdout states retain the current rating grade within the one-year horizon. Thus, knowing that a state survives to \( T^* = 0.25 \) increases the probability that it is a survivor. As a result, survival estimates conditional on survival time \( T^* = 0.25 \) year exhibit better predictive accuracy than the unconditional survival estimates. Short-term forecasts conditional on \( T^* = 0.25 \) year are less optimistic and achieve smaller bias than unconditional forecasts.

7.4. Conclusion

This chapter extends the work of chapters 5 and 6, and attempts to investigate the predictive accuracy of the probability survival forecast generated by the TVC base model and TVC extended model. The forecast accuracy differs between downgrades and upgrades. The Brier scores and their decomposition indicate that the models’ forecasts had some predictive power and showed some forecasting skill. However, considered overall, it cannot be claimed that the models beat the prior frequency benchmark. The models may have had an edge at shorter horizons, but this was lost at longer horizons, particularly for forecasts extending for more than two years.

The somewhat disappointing forecast performance of the models may in part be due to a very demanding test, since the credit conditions in the holdout period changed so dramatically relative to the estimation period. The model may also have been
handicapped by the employment in Equation (9) of static macro-economic data for holdout states. Implementing forecasts in models with time-varying variables is not trivial and one avenue for further research is to use forecast values of the time-varying variables. It also needs to be remembered that these models were not designed with the intention of forecasting; rather the intention was to examine the impact of rating history and time on rating migrations. The use of these variables with other variables might well yield a model with greater predictive accuracy. For instance, the introduction of CreditWatch or rating outlook data as a time-varying variable would probably improve the forecast performance of the TVC hazard models36.

Examining the comparative forecast performance of the stratified TVC base models and the stratified time-fixed models, the following conclusions are drawn. For downgrades, accounting for the macro-economic changes over time improves the predictive accuracy of the model at the short-term horizon. However, the TVC base model, compared to the time-fixed model, exhibits more bias and performs poorly at the longer-term horizon, where it generates survival estimates that are more pessimistic than those formed by the time-fixed model. For upgrades, relative to the time-fixed model, the TVC base model achieves comparable forecast performance at the short-term horizon. The improvement of the upgrade model is most evident at the longer forecast horizon (three to five-year lead time). The dynamic estimation framework, which captures macro-economic conditions over time, generates lower upgrade estimates and higher survival estimates during downturns. The TVC model is

36 Hamilton and Cantor (2004) found that rating history and rating outlooks are each “highly predictive” of subsequent rating migrations, however, rating history displays “little additional impact” on rating changes once outlook status were taken into account. Vazza, Leung, Alsati, and Katz (2005, p. 1) demonstrated that “CreditWatch status and outlooks are strong predictor of rating behaviour” after controlling for important factors such as initial ratings.
therefore able to produce longer-term survival estimates, which are more optimistic than those generated by the time-fixed model and reflects the long-lived nature of up states. However, both upgrade models display poor discrimination ability.

The TVC extended model allows for the interaction between time and rating history. Forecasts can only be made conditional on holdout states surviving to the time \( T^* \) at which point the interaction terms are constructed. These survival estimates are therefore referred to as conditional survival probabilities. Adding the interaction covariates between rating history and time introduces two effects to the conditional survival estimates, the decay effect and the information advantage effect. The decay effect is driven by the extension of time, whereas the information advantage effect is driven by the changing sample composition as time extends and issuers leave the sample.

For downgrades, conditional survival estimates are higher, more optimistic in the short term and less pessimistic in the longer term, than the unconditional survival estimates generated by the TVC base model. Short-term survival forecasts conditional on survival time \( T^* = 0.25 \) year have a larger positive bias. The improvement of the TVC downgrade model after accounting for time is most evident at the longer-term horizon as the conditional survival time \( T^* \) extends from a quarter to a year. The conditional survival time \( T^* = 1 \) year conveys useful information as holdout states surviving the first year are likely to have lower one-year downgrade probabilities. The presence of the interaction terms constructed at \( T^* = 1 \) year brings about an information advantage that helps diminish bias. Of particular interest, the two-year forecasts conditional on issuers surviving at \( T^* = 1 \) year have a minimal bias.

For upgrades, conditional survival estimates are less optimistic in the short term and more pessimistic in the longer term, than unconditional survival estimates. The
improvement of the TVC upgrade model after accounting for time is most evident at the short-term horizon. Forecasts conditional on surviving time $T^* = 0.25$ year have less positive bias and better discrimination than the unconditional forecasts generated by the TVC base model. This can be attributed to the information advantage embedded in the significant interaction terms between rating history and the conditional time $T^*$. However, the advantage of the TVC extended model does not hold at longer-term forecast horizons when the conditional survival time $T^*$ is extended to one year or longer. It seems that the decay effects become pronounced as time extends by yearly intervals. This translates into lower (more pessimistic) survival estimates and a larger negative bias.

In conclusion, the dynamic estimation framework that captures the evolution of macro-economic conditions over time improves the forecasting performance of the stratified hazard models, at the short-term horizon for downgrades and at the longer-term horizon for upgrades. Adding the interaction terms between rating history and survival time improves the predictive accuracy of the stratified TVC hazard models. The relative performance varies between downgrades and upgrades, and depends on how informative the conditional survival time $T^*$ is.
Chapter 8

Conclusion

This chapter first draws conclusions from the research carried out for this thesis, then discusses the contributions and implications of the empirical results. The chapter closes with a discussion of the limitations and possible directions for future research.

8.1. Summary of findings

Despite being the most popular method used by credit rating agencies to derive a distribution for credit rating migrations, the discrete time cohort Markov framework is not strongly supported by empirical studies. The main underlying assumption is that the rating process is Markov and time-homogeneous. Consequently, this framework ignores the heterogeneity across issuers of the same rating grade and the time-heterogeneity in rating dynamics. In addition, the migration matrix is constructed from the relative frequencies of rating migrations observed in the past. The framework therefore fails to provide non-zero probability estimates of rare or unobserved events which leave no historical record, for instance a migration from AAA to the default state.

In empirical studies, the qualitative response static models have been widely employed as alternatives to the discrete time cohort Markov framework. These static models allow an estimate that an issuer, given its characteristic at a given point in time (static), will survive or experience a migration event. However, the static models fail to account for changes in risk factors over time, and do not provide a forecast as a function of time. The latter limitation poses critical issues for banks in calculating time-varying economic risk capital requirements, and in detecting deterioration in the credit quality of investment portfolios within a sufficient lead time.
In addition, both the discrete time cohort Markov framework and the qualitative response static models suffer from other critical issues. Both frameworks ignore the migration times and the survival durations of issuers in the sample. Information pertaining to issuers that leave the study for reasons other than the migration event is not utilised in the estimation process, resulting in an efficiency loss.

Since the new Basel Accord framework came into effect, there has been ongoing interest in estimating time-varying rating migration probability for both risk management and capital adequacy purposes. Modelling rating migration is facilitated by an understanding of the underlying rating migration dynamics. The literature has provided ample evidence that rating history impacts on rating distribution and rating migrations vary over time. This has motivated academic studies to continue looking for alternative estimation approaches that account for the time-heterogeneity and the issuer-heterogeneity in rating dynamics, and that solve the issues faced by the conventional approaches.

8.1.1. Estimated models

This thesis employs Cox’s hazard model (Cox, 1972) to estimate the probability that a rating survives in its current grade. The Cox’s hazard model overcomes the critical problems discussed above. It allows a rigorous testing of non-Markovian behaviours in rating dynamics, and provides estimates of the probability of survival through time. Importantly, it is particularly convenient to introduce a time dimension into the risk factors examined and to estimate dynamic Cox’s hazard model (with time-varying covariates).

The conditional gap time estimation approach applied in this thesis specifically controls the time sequence of recurrent events (Prentice et al., 1981). The Cox’s
hazard models were developed as stratified models; consequently, observations of the same migration sequence (stratum) share one distinct baseline hazard function. The thesis develops not only a stratified Cox’s proportional hazard model (conventional time-fixed model) but also two stratified dynamic Cox’s hazard models with time-varying covariates (TVC base and TVC extended models) to investigate non-Markovian behaviours in rating dynamics. The earlier model employs a static estimation framework, whereas the latter models apply a dynamic estimation framework that accounts for the changing macro-economic conditions over time. Of the two dynamic models, the stratified TVC extended model further controls for the interaction between time spent in a rating grade and the main effect variables.

The empirical studies in chapter 5 (time-fixed model) and chapter 6 (TVC models) examine the impact of a number of aspects of rating history on migration hazard after controlling for industry heterogeneity, the business cycle and duration dependence. In general, the findings of the time-fixed and TVC models offer improved understanding of non-Markovian behaviour in rating dynamics over the period 1984-2000. The results are consistent with the evidence found in previous studies and as hypothesised in chapter 2. First, future rating changes depend not only on the current rating (start rating) but also on a substantial number of past rating behaviours. Downgrades show more dependence on rating history than upgrades. Second, some aspects of rating history exhibit the tendency to repeat themselves. Issuers with longer lagged durations (lag one, lag two) have a higher probability of continuing in the current rating state. Issuers with a volatile downgrade history (rate prior down), or a downgrade at lag one rating state (dummy lag1 down) are more like to be downgraded and less likely to be upgraded. Third, different rating paths lead to different rating distributions. A fallen angel (number FA) is more likely to ascend and less likely to descend the rating
spectrum. A break in rating history by being unrated (number NR) makes a further
downgrade less likely. Fourth, considered individually some rating history covariates
are more influential on subsequent rating changes than the current rating (start rating).
For instance, lagged rating changes have a particularly large effect, increasing the
hazard of a rating migration by hundreds of percentage points. A downgrade at lag
one rating state (dummy lag1 down) makes a downgrade 157 percent more likely in a
static framework (time-fixed model), and 149 percent more likely in a dynamic
framework without accounting for time (TVC base model). Fifth, the current rating
(start rating) has a relatively small impact. A higher rating increases the probability
that the current rating persists. Furthermore, being rated in the boundary of investment-
grade ratings (dummy inv boundary) or speculative-grade ratings (dummy junk
boundary) diminishes the probability of a downgrade but raises the chance of an
upgrade. Sixth, the effect of rating history on subsequent rating changes is strong and
persists in both the static estimation framework (time-fixed model) and the dynamic
estimation framework (TVC base models). In other words, rating dynamics exhibit
non-Markovian behaviour, and the impact of rating history remains strong after
accounting for the evolution of macro-economic conditions over time.

This thesis provides new evidence on the interactions between rating history and the
time for which the current rating persists. The results of the stratified TVC extended
hazard models estimated for the period 1984-2000 (chapter 6) show that the impact of
rating history and current rating (start rating) variables diminishes as the rating
duration gets longer. In other words, some aspects of rating history and the current
rating (start rating) interact with time, and the time interactions reduce the main
effects of rating history variables on migration hazard. This decay effect is
particularly strong in the cases of issuers with a downgrade at lag one rating state
(dummy lag1 down). Without controlling for the passage of time, these issuers are estimated to be 149 percent more likely to travel down the rating spectrum. After accounting for interactions with time, the 149 percent increases to 309 percent at the start of the rating. However, the risk decreases as the current rating state persists through time, and drops to 215 percent at the end of the first year. This result implies that failing to account for the decay of the main effects arising from the interactions between rating history and time could result in substantial errors in estimating migration probabilities.

8.1.2. Forecast accuracy assessment

A test of the models’ predictive ability using a holdout sample (2001-2005) was conducted. Based on the stratified time-fixed hazard model (chapter 5), the stratified TVC base and TVC extended hazard models (chapter 6) estimated for the period 1984-2000, time-varying probabilities of survival were estimated for the subsequent period, 2001-2005. This provided a considerable challenge for the model, as the rating conditions changed dramatically between the estimation and holdout periods. By employing the Brier score (Brier, 1950) and its covariance decomposition (Yates, 1982), the forecast performance of quarterly estimates within a one-year window and yearly estimates within a five-year window were assessed. The Brier score skill components (bias, scatter, and slope) of probability survival estimates generated by each estimated hazard model at a specific horizon were then compared to the relevant skill components of the benchmark Brier score formed at the same lead time.

Both downgrade and upgrade models show poor discrimination ability. In the short term, the downgrade models perform poorly, except at three-quarter and one-year forecast horizons. The underperformance is due to a strong tendency to place short-term survival forecasts in the probability categories substantially above the survival
base rate. This propensity translates into a large (positive) bias in short-term survival estimates. On the other hand, the upgrade models show modest ability in matching mean short-term survival forecasts to mean relative frequency, and offer well-calibrated estimates at one and two-year time horizons. This tendency corresponds to a small (positive) bias within two-year forecast horizons. Both the downgrade and upgrade models assign longer-term survival estimates to the probability categories substantially below the observed relative frequency. Consequently, longer-term estimates have a large (negative) bias.

In the static estimation framework, the forecast performance of the proportional downgrade model is generally disappointing, whereas the proportional upgrade model exhibits modest predictive accuracy within two-year horizons.

Controlling for the changes in the macro-economic environment over time using the TVC base model slightly improves the forecast performance of the downgrade model within a one-year forecast horizon and substantially improves the predictive accuracy of the upgrade model at three- to five-year forecast horizons.

The TVC extended model allows for the interaction between time and rating history. In this context, forecasts can only be made conditional on survival to the date at which the interaction terms are computed. These estimates are therefore referred to as conditional survival probabilities. The presence of the time interaction term introduces two effects on the conditional survival probabilities. These effects are the information advantage effect and the decay effect. The information advantage effect is driven by the changing sample composition as time extends and issuers with volatile rating history leave the sample. On the other hand, the decay effect is driven by the extension of time, and is embedded in the interaction terms. Both the magnitude and the sign of the main rating history variables influence the decay effect. Forecast
performance is driven by both the extension of time (the decay effect) and the changing sample composition (the information advantage effect).

Accounting for the time interactions substantially improves the forecast performance of conditional survival estimates generated by the TVC extended downgrade model at the longer-term horizon, and slightly improves the predictive accuracy of conditional survival estimates formed by the TVC extended upgrade model at the short-term horizon.

8.2. Contribution and implications

The findings presented in this thesis contribute to the research literature and have some practical implications.

8.2.1. Contribution

The contribution of this thesis is fourfold.

First, the findings strengthen and extend the evidence of non-Markovian behaviours in rating dynamics. The empirical studies carried out for this thesis examine the question of rating history dependence and time dependence in the rating migration process after controlling for macro-economic conditions and industry-heterogeneity. The estimated models incorporate well-documented empirical properties in rating migration dynamics such as the current rating (\textit{start rating}), downward momentum (\textit{dummy lag1 down, dummy lag2 down}), duration dependence (\textit{lag one, lag two}), fallen angel event (\textit{number FA}), rating age (\textit{age since first rated}), and the first rating (\textit{original rating}).

The thesis presents new evidence on additional aspects of rating history and the current rating, which received little attention in previous studies, such as the proximity of the current rating to the boundary of investment / speculative rating
grades (dummy inv boundary, dummy junk boundary), rating change volatility (rate prior change), downgrade volatility (rate prior down), rating withdrawals (number NR), prior multiple-notch downgrades and upgrades (number big down, number big up), and the occurrences of rising star events (number RS). The thesis contributes to the literature by offering an improved understanding of non-Markovian behaviours observed in both static and dynamic estimation frameworks. A key finding is that past rating behaviours persist even after accounting for the dynamic development of the macro-economic environment over time.

Second, the thesis makes a contribution to the literature by offering new evidence that the effect of rating history on migration probability interacts with time and decays as time extends. As the current rating persists, the distance in time from past rating behaviour extends. It is not surprising that the extension of time makes the more distant variables less relevant. The argument is supported by the results that the duration (lag two) and a downgrade at lag two rating state (dummy lag2 down) have smaller effects on migration hazard than the duration (lag one) and a downgrade at lag one rating state (dummy lag1 down). In the same vein, rating age (age since first rated) and the first rating (original rating) have minimal impact relative to other rating history variables. However, the key finding is that as the time spent in the current rating gets longer, rating history variables interact with the survival duration of the current rating state. The resulting decay effects are substantial. The findings suggest that failure to account for the time interaction phenomenon can lead to substantial errors in the estimation of migration risk at a point in time.

Third, the thesis contributes to the framework for estimating migration models by applying an innovative approach, the stratified dynamic Cox’s hazard model (Cox, 1972) with time-varying covariates. The power and the flexibility of Cox’s hazard model are enhanced by incorporating time-varying covariates. The stratified dynamic Cox’s hazard model accounts for the dynamic nature of rating changes and the persistence of rating behaviours. The model is particularly useful for understanding the complex interplay of rating changes and their effects on migration probabilities. By incorporating time-varying covariates, the model provides a more accurate estimation of migration probabilities, taking into account the dynamic nature of rating changes and the persistence of rating behaviours.
model makes it well suited to account for recurrent migration events, to control for the changes in macro-economic covariates over time, and to account for the passage of survival time over rating durations. The thesis overcomes the challenges in estimating stratum-specific baseline hazard function and forming time-varying probability estimates when the proportionality assumption does not hold and the data sample includes multiple strata.

Fourth, this thesis enriches the literature on the framework for evaluating forecasts of migration probabilities by employing the Brier score (Brier, 1950) and its Yates covariance decomposition (Yates, 1982). This approach has received little attention in financial studies, though it has been widely applied in meteorology. The approach offers the possibility to evaluate forecast accuracy in terms of discrimination, calibration and scatter. It provides insights into forecast characteristics and sources of forecast errors.

8.2.2. Implications

The findings of the research carried out for this thesis are relevant for fixed-income portfolio managers, banking institutions and regulators. There are also implications for portfolio investment and credit risk management.

Knowledge of past rating behaviours is useful in estimating the investment performance of active managed portfolios and in establishing the portfolio drift tolerance level (Altman and Kao, 1991; Altman and Kao, 1992b). Given the evidence of strong rating momentum, an initial rating change can provide a signal about future investment performance.

The Basel II framework institutes rules that banks map each credit rating onto a capital charge. The behaviours of rating history can signal changes in economic
capital requirements. Consequently, past rating behaviours provide relevant information to the establishment of loss reserves and to allocating risk capital efficiently (Altman and Kao, 1991).

The strong effect of rating history on migration hazard explicitly rules out the Markov property, which is commonly assumed in the discrete time cohort method and widely applied by credit rating agencies to estimate the migration matrix. The point is that internal rating-based models should incorporate rating history variables to account for the path that an issuer has followed to the current rating state. Separate hazard models could be developed to account for the varying risk of rating changes experienced by issuers with different historic rating paths, and to capture the different rating process of downgrades and upgrades. Where point in time estimates of migration probabilities are required, the interaction between time in the current rating and rating history needs to be considered. Failure to do this may well result in substantial estimation errors.

It is suggested that the estimation method matters statistically and economically when estimating the migration matrix and that there are efficiency gains in using duration approaches to estimate the required risk capital (Jafry and Schuermann, 2004). The estimation approach proposed in this thesis, the stratified dynamic Cox’s hazard model (Cox, 1972), not only allows for the duration at risk of each issuer but also captures the phenomena of rating history dependence and time dependence. It therefore has advantages over the discrete time cohort Markov framework and the conventional static models. The technical framework proposed in this thesis can be employed by banks to estimate hazard models that are predictive of rating changes. The appealing feature of generating time-varying forecasts offers banks the possibility
of taking preventive or remedial action, and indicating how quickly this needs to be done.

The methodological framework presented in this thesis for evaluating forecast performance, the Brier score (Brier, 1950) and its covariance decomposition (Yates, 1982) can be employed for the assessment of the internal rating system applied by banks subject to the Basel II regulations. The simple calculation, the intuitive nature of covariance components, and the informative graphical depictions make the Brier score’s covariance decomposition a useful tool to assess the predictive accuracy of probability estimates. Furthermore, the decomposition provides useful feedback for banks to refine and improve credit risk models and credit rating systems.

8.3. Limitations and further research

The empirical studies conducted in this thesis suffer from some limitations and suggest directions for future research.

First, changes in the rating outlook or CreditWatch listings of issuers were not incorporated into the models, as the credit rating outlook information was not available in the ratings database used for this study. The use of rating outlook or CreditWatch data as a time-varying variable in the TVC hazard models might diminish the impact of the rating history variables (Hamilton and Cantor, 2004) and might also lead to improved forecast performance (Vazza, Leung, Alsati, and Katz, 2005).

Second, the numeric coding of rating spectrum from 0 (Default) to 26 (AAA) contains more rating notches than Standard & Poor’s alphabetical rating scale. The numeric conversion assumes multiple notches between the upper-end (AAA) and lower-end (CC, C, D) and the rest of the rating spectrum. It might be suggested that this has an impact
on the results and the measurement of big down jumpers/ big up jumpers. However a substantive impact seems unlikely as very few issuers (2.24 percent of the estimation sample) were rated at the top and the bottom of the rating scales (see Table 4-7). Furthermore, a very small number of states in the estimation sample either departed from or ended at AAA, CC, C (see Table 4-3B), and few cases were treated as big rating jumpers owing to the multiple notch rating scale. Table 8.1 shows that if the numeric rating codes exactly matched Standard and Poor’s rating scale (i.e. rating AAA coded 25 instead of 26) then the 16 of the downgrades and 4 of the upgrades shown in Table 4-7 would not be considered big jumpers.\(^{37}\)

Table 8-1

<table>
<thead>
<tr>
<th>Migrations</th>
<th>Big down jumpers</th>
<th>Big up jumpers</th>
</tr>
</thead>
<tbody>
<tr>
<td>From CC to C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>From CCC to CC</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>From AAA to A</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>From C to CC</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>From AA to AAA</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>

Third, the estimated downgrade models do not fit well in the holdout period. The empirical frequency distribution is more robust than the model forecasts in longer-term horizon. Several factors could account for this result. The models were estimated over a long period of time, 1984-2000, and cyclical effects were smoothed out. The estimation period is clearly not representative of the volatile holdout period. Furthermore, the employment of static macro-economic data for holdout states (see Figure 4-4) in Equation (9) is likely to dampen the predictive accuracy of the

\(^{37}\) A substantial rating change is a downgrade of at least three rating notches and an upgrade of at least two rating notches.
estimated models, given the dramatically changing macro-economic conditions in the holdout period, 2001-2005.

A natural extension of the study, therefore, would be to continually update the model by using a moving window and/or recalibrating the model as changes in rates of migration become evident. Regularly updating the macro-economic covariates \( Z^q(t) \) used in Equation (9) to form estimates for holdout state \( q \) by using time-series forecasts is also a possible direction for future research. The employment of dynamic macro-economic data in the form of time-series forecasts for holdout observations would control for the expected changes of macro-economic conditions over the holdout period, and would introduce a forward-looking perspective into the survival forecasts.

The research could also be extended by estimating two additional versions of the hazard models, one for economic contraction periods, and one for economic expansion periods. This approach allows the examination of credit quality changes under both favourable and unfavourable macro-economic conditions. The application of either the contraction model or the expansion model to generate survival probability forecasts would then depend on the forecast of macro-economic conditions updated regularly during the holdout period. A more rigorous examination of this issue is left for future research.

Fourth, in forming \( t \)-period survival forecasts, rating states that had either been censored or had experienced the event prior to the forecast time \( t \) were excluded from the holdout sample. The use of \( \textit{ex-post} \) information in generating forecasts points to a possible weakness of Chen \textit{et al.} (2005)’s approach applied in this thesis. This approach results in a changing composition of the holdout sample as the forecast
period unfolds. Consequently, forecast performance at different horizons was driven
by both time and the changing sample. Future work could be directed to forecasts
where the sample composition is held constant as the forecast horizon extends.
References


Lawrence Edbaum Associates. Mahweh, NJ.


Johnson, R. 2004. Rating agency actions around the investment grade boundary. 


Appendix A: The two sample t test and the Wilcoxon two sample test

| Variable                                      | Sample size | Mean    | Median   | Sample size | Mean    | Median   | Method       | Variance | t statistic | Pr > |t| Z statistic | Two-Sided Pr > |Z| |
|-----------------------------------------------|-------------|---------|----------|-------------|---------|----------|--------------|----------|-------------|-------|-------------|----------------|-------|-------|
| CFNAI                                         | 68          | 0.113   | 0.155    | 20          | -0.301  | -0.353   | Pooled Equal | -3.6     | 0.0005      | -3.24 | 0.0012      |                |       |       |
| RealGDPg actual minus potential               | 68          | 0.131   | 0.153    | 20          | -0.160  | -0.194   | Pooled Equal | -3.28    | 0.0015      | -2.87 | 0.0041      |                |       |       |
| Industrial production change                  | 68          | 0.939   | 1.011    | 20          | 0.183   | 0.420    | Satterthwaite Unequal | -4.5     | <.0001  | -3.92 | <.0001       |                |       |       |
| SP500 quarterly return                        | 68          | 3.609   | 3.294    | 20          | -0.501  | 0.157    | Pooled Equal | 3.22     | 0.0018      | 2.93  | 0.0034      |                |       |       |
| SP500 annual SD                               | 68          | 1.769   | 1.678    | 20          | 2.284   | 2.467    | Satterthwaite Unequal | 3.68     | 0.0002      | 3.68  | 0.0002      |                |       |       |
| Term structure slope                          | 68          | 1.716   | 1.718    | 20          | 2.140   | 2.621    | Pooled Equal | 1.73     | 0.0869      | 1.94  | 0.0516      |                |       |       |

Table A-2: Proportion of upgrades/ downgrades

| Dummy up | 4487 | 0.2619 | 0 | 1872 | 0.1309 | 0 | Satterthwaite Unequal | -12.85 | <.0001 | -11.4306 | <.0001 |
| Dummy down | 4487 | 0.3784 | 0 | 1872 | 0.4279 | 0 | Pooled Equal | 3.68 | 0.0002 | 3.68 | 0.0002 |

Table A-3: Time to event (survival duration)

| Survival duration, all states | 4487 | 2.1288 | 1.4164 | 1872 | 0.9441 | 0.6425 | Satterthwaite Unequal | -29.53 | <.0001 | -22.8682 | <.0001 |
| Survival duration, down states | 1698 | 1.658  | 0.997  | 801  | 0.514  | 0.296  | Satterthwaite Unequal | -22.22 | <.0001 | -18.0515 | <.0001 |

The F-statistic (not reported) is used to decide whether the hypothesis that the two populations have equal variances can be rejected at 5 percent significance level.

* Dummy up is a dummy variable that takes the value of one if the state is an up state, and zero otherwise. The mean of variable dummy up equal to the proportion of upgrades in the sample.

** Dummy down is a dummy variable that takes the value of one if the state is a down state, and zero otherwise. The mean of variable dummy down equal to the proportion of downgrades in the sample.
**Appendix A: The two sample t test and the Wilcoxon two sample test (Continued)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimation sample</th>
<th>Holdout sample</th>
<th>T test</th>
<th>Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td>Mean</td>
<td>Median</td>
<td>Sample size</td>
</tr>
<tr>
<td>Start rating, all states</td>
<td>4487</td>
<td>15.610</td>
<td>16 (BBB-)</td>
<td>1872</td>
</tr>
<tr>
<td>Start rating, down states</td>
<td>1698</td>
<td>14.715</td>
<td>16 (BBB-)</td>
<td>801</td>
</tr>
<tr>
<td>Start rating, up states</td>
<td>1175</td>
<td>15.570</td>
<td>16 (BBB-)</td>
<td>245</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimation sample</th>
<th>Holdout sample</th>
<th>T test</th>
<th>Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td>Mean</td>
<td>Median</td>
<td>Sample size</td>
</tr>
<tr>
<td>Age since first rated (year)</td>
<td>4487</td>
<td>8.578</td>
<td>7.7342</td>
<td>1872</td>
</tr>
<tr>
<td>Original rating</td>
<td>4487</td>
<td>17.254</td>
<td>17 (BBB)</td>
<td>1872</td>
</tr>
<tr>
<td>Lag one (year)</td>
<td>4487</td>
<td>2.131</td>
<td>1.5343</td>
<td>1872</td>
</tr>
<tr>
<td>Dummy lag1 down</td>
<td>4487</td>
<td>0.600</td>
<td>1</td>
<td>1872</td>
</tr>
<tr>
<td>Lag two (year)</td>
<td>4487</td>
<td>2.183</td>
<td>1.6274</td>
<td>1872</td>
</tr>
<tr>
<td>Dummy lag2 down</td>
<td>4487</td>
<td>0.600</td>
<td>1</td>
<td>1872</td>
</tr>
<tr>
<td>Rate prior change</td>
<td>4487</td>
<td>0.640</td>
<td>0.5143</td>
<td>1872</td>
</tr>
<tr>
<td>Rate prior down</td>
<td>4487</td>
<td>0.413</td>
<td>0.2841</td>
<td>1872</td>
</tr>
<tr>
<td>Number NR</td>
<td>4487</td>
<td>0.077</td>
<td>0</td>
<td>1872</td>
</tr>
<tr>
<td>Number FA</td>
<td>4487</td>
<td>0.262</td>
<td>0</td>
<td>1872</td>
</tr>
<tr>
<td>Number RS</td>
<td>4487</td>
<td>0.199</td>
<td>0</td>
<td>1872</td>
</tr>
<tr>
<td>Number big down</td>
<td>4487</td>
<td>0.301</td>
<td>0</td>
<td>1872</td>
</tr>
<tr>
<td>Number big up</td>
<td>4487</td>
<td>0.394</td>
<td>0</td>
<td>1872</td>
</tr>
</tbody>
</table>

* The F-statistic (not reported) is used to decide whether the hypothesis that the two populations have equal variances can be rejected at 5 percent significance level.
Appendix A: The two sample t test and the Wilcoxon two sample test (Continued)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimation sample</th>
<th>Holdout sample</th>
<th>T test</th>
<th>Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size</td>
<td>Mean</td>
<td>Median</td>
<td>Sample size</td>
</tr>
<tr>
<td>FA*</td>
<td>4487</td>
<td>0.2405</td>
<td>0</td>
<td>1872</td>
</tr>
<tr>
<td>RS**</td>
<td>4487</td>
<td>0.1859</td>
<td>0</td>
<td>1872</td>
</tr>
<tr>
<td>Dummy lag1 down</td>
<td>4487</td>
<td>0.6002</td>
<td>1</td>
<td>1872</td>
</tr>
<tr>
<td>Dummy lag2 down</td>
<td>4487</td>
<td>0.5995</td>
<td>1</td>
<td>1872</td>
</tr>
</tbody>
</table>

*The F-statistic (not reported) is used to decide whether the hypothesis that the two populations have equal variances can be rejected at 5 percent significance level.

*FA* is a dummy variable that takes the value of one if the issuer experienced one or more fallen angel events in the past, and zero otherwise. The mean of variable FA equal to the proportion of issuers which experienced fallen angel event(s) prior to the current rating state.

**RS** is a dummy variable that takes the value of one if the issuer experienced one or more rising star events in the past, and zero otherwise. The mean of variable RS equal to the proportion of issuers which experienced rising star event(s) prior to the current rating state.