Abstract

He, Xiaofeng. Credit Cycle, Credit Risk and Business Conditions (Under the direction of Professor Peter Bloomfield.)

We first present a Complex Singular Value Decomposition (CSVD) analysis of credit cycle and explore the lead-lag relation between credit cycle and business cycle, then propose a Generalized Linear Model (GLM) of credit rating transition probabilities under the impact of business conditions.

To detect the cyclic trend existence of credit condition in U.S. economy, all credit variables and business variables are transformed to complex values and the transformed data matrix is approximated by first order of CSVD analysis. We show that the economy, represented by both credit conditions and business conditions, is changing recurrently but with different frequencies for different time periods. Credit variables making the greatest linear contribution to first Principal Component can be identified as credit cycle indicators. The result of leading business variables to credit variables in an economy provides the basis to predict credit condition by business cycle indicators.

The credit rating system is a publicly available measure of the riskiness of financial securities and a rating transition matrix quantifies the risk, by permitting calculation of the probability of downgrade or default. Credit migration is observed to be influenced both by business conditions and by an issuer’s own credit status. We assume the rating history for a particular institution is Markovian, and histories for different institutions are assumed to be statistically independent, in both cases
the history of market conditions are known. With a simple GLM, we investigate the significance of business conditions and their two major impacts - creditworthiness deterioration/improvement and credit stability. We propose a model of transition probability in discrete time and a model of instantaneous transition rates in continuous time, and fit them by maximum likelihood. Business conditions are shown to have a significant effect: higher likelihood for credit quality improvement and stability under good business conditions while higher likelihood for credit quality deterioration and drift under severe business conditions. The two business impacts are significant and business deterioration/improvement impact is greater than its stability impact on credit rating transitions. Investment-grade rating transitions are more sensitive to long rate risk while speculative-grade rating transitions are more sensitive to short rate risk. Compared to a discrete model, the continuous transition model has much greater over-dispersion but is more practical.
CREDIT CYCLE, CREDIT RISK AND BUSINESS CONDITIONS

by

XIAOFENG HE

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

STATISTICS

in the

GRADUATE SCHOOL

at

NC STATE UNIVERSITY

2001

Professor P. Bloomfield, Chair

Professor J.P. Fouque

Professor S. G. Pantula

Professor D. A. Dickey
To my parents and my husband
Biography

Xiaofeng He was born in Tianjin, China to parents Changwen He and Guirong Yu on November 1st, 1972. She was awarded a B.S. degree in Information Science from Nankai University, China in June 1995. Since then she had been with Tianjin NEC Co. in China before she came to U.S. in August, 1996. She received her M.S. degree in Statistics from North Carolina State University in May, 1998. After that, she continued to work towards the Ph.D degree in Statistics at North Carolina State University, under the direction of Dr. Peter Bloomfield.
**Acknowledgements**

I am deeply indebted to my doctoral advisor, Professor Peter Bloomfield. Without his continuous encouragement and resourceful ideas, my study will never turn into this final dissertation. Beyond this research, he taught me how to warmly embrace failure, how to open my mind, and how to believe in myself. Having been his student for the past three years is a valuable experience which will definitely benefit my future researches.

I am grateful to Prof. Jean-Pierre Fouque for teaching me so many subjects in mathematical financial engineering field. I admire his active and enthusiastic involvement in both academic researches and financial industry applications. I also would like to thank Prof. Dickey and Prof. Pantula for their support, feedback and comments on my dissertation.

Thank you, my wonderful husband and my best friend, Jingyu Sun. He contributes to my research by all means, his computer talent, his patient understanding, and his love during the course of this study.

Finally, my deepest gratitude goes to my parents. Their unconditional love and support have made who I am today. I hope they can be proud of my doctorate fulfillment.
Contents

List of Figures viii
List of Tables ix

I A Study on Credit Cycle and Business Cycle 1

1 Introduction 2
  1.1 Motivation 2
  1.2 Data description 3
    1.2.1 Business variables 4
    1.2.2 Credit variables 5
    1.2.3 Credit distressed variables 6
    1.2.4 Data preparation 7

2 Complex Singular Value Decomposition Analysis 8
  2.1 Complex value transformation 9
  2.2 Complex singular value decomposition analysis 14
  2.3 First order complex singular value decomposition result 18
  2.4 Summary 23

II A Generalized Linear Model of Business Condition on Credit Risk Drift Process 30

3 Introduction 31
  3.1 Background 31
  3.2 Literature review on modeling credit risk 38
  3.3 Research outline for Part II 44
4 Data and Assumptions
  4.1 Data description ........................................ 46
    4.1.1 Credit rating transition matrix ....................... 47
    4.1.2 Macro-economic time series .................................. 50
  4.2 Fundamental assumptions ........................................ 54
    4.2.1 Assumption 1 - identical transition distribution for issuers with same initial rating ....................... 54
    4.2.2 Assumption 2 - business dependence of transition distribution .................................. 55
    4.2.3 Assumption 3 - single-issuer Markov rating transition process .................................. 56
    4.2.4 Assumption 4 - multi-issuer independent transition distributions .................................. 58
    4.2.5 Consideration of business conditions ....................... 60

5 Discrete Transition Probability Models ........................................ 62
  5.1 One-year discrete transition models .................................. 63
    5.1.1 Null transition model .................................. 67
    5.1.2 One-covariate transition model ....................... 68
    5.1.3 Two-covariate transition model ....................... 68
  5.2 Maximum likelihood estimation .................................. 69
    5.2.1 Likelihood function .................................. 69
    5.2.2 Maximum likelihood estimation implementation ....................... 75
  5.3 Goodness-of-fit test, over-dispersion and scaled-deviance .............. 80
    5.3.1 Goodness-of-fit \( \chi^2 \) test .................................. 80
    5.3.2 Over-dispersion .................................. 81
    5.3.3 Scaled-deviance .................................. 83
  5.4 Observed variance of maximum likelihood estimate ....................... 86

6 Continuous Transition Probability Models .................................. 90
  6.1 Data preparation .................................. 92
    6.1.1 Square transition matrix modification ....................... 92
    6.1.2 Monthly business time series modification ....................... 93
  6.2 Continuous transition probability models .................................. 93
    6.2.1 Derivation of the constant instantaneous transition matrix ....................... 93
    6.2.2 Instantaneous transition probability models .................................. 97
      6.2.2.1 Null instantaneous transition model ....................... 99
      6.2.2.2 One-covariate instantaneous transition model ....................... 99
      6.2.2.3 Two-covariate instantaneous transition model ....................... 100
  6.3 Maximum likelihood estimation and optimization implementation ....................... 101
    6.3.1 Likelihood function .................................. 101
    6.3.2 Powell’s quadratic numerical optimization ....................... 103
  6.4 Goodness-of-fit test, over-dispersion and scaled-deviance ....................... 104
7 Part II Results

7.1 Results for discrete transition models
   7.1.1 Complete-grade transitions
      7.1.1.1 Null model
      7.1.1.2 The one-covariate model
      7.1.1.3 Two-covariate model
   7.1.2 Investment-grade transitions
      7.1.2.1 Null model
      7.1.2.2 One-covariate model
      7.1.2.3 Two-covariate model
   7.1.3 Speculative-grade transitions
      7.1.3.1 Null model
      7.1.3.2 One-covariate model
      7.1.3.3 Two-covariate model
   7.2 Results for continuous transition models
      7.2.1 Null model
      7.2.2 One-covariate model
      7.2.3 Two-covariate model
   7.3 Conclusion

8 Summary and Future Work

8.1 Summary
8.2 Future work

Bibliography
List of Figures

2.1 Amplitude of first score vector .......................................... 24
2.2 Phase time series of the first CPC ....................................... 25
2.3 Frequency time series of the first CPC .................................. 26
2.4 Period time series of the first CPC ...................................... 27
2.5 Loadings of the first CPC ................................................... 28
2.6 First CPC representation of one economy cycle ...................... 29
## List of Tables

4.1 S&P 1999 rating transition matrix ........................................... 47  
7.1 Estimated one-year transition matrix under the discrete null model . 130  
7.2 Discrete one-covariate stability-migration model results for complete-grade transitions ................................................................. 131  
7.3 Discrete one-covariate upgrade-downgrade model results for complete-grade ................................................................. 132  
7.4 Top ten best discrete two-covariate model results for complete-grade transitions ................................................................. 133  
7.5 Discrete one-covariate stability-migration model results for investment-grade transitions ................................................................. 134  
7.6 Discrete one-covariate upgrade-downgrade model results for investment-grade transitions ................................................................. 135  
7.7 Top ten best two-covariate model results for investment-grade transitions ................................................................. 136  
7.8 Discrete one-covariate stability-migration model results for speculative-grade transitions ................................................................. 137  
7.9 Discrete one-covariate upgrade-downgrade model results for speculative-grade transitions ................................................................. 138  
7.10 Top ten best discrete two-covariate model results for speculative-grade transitions ................................................................. 139  
7.11 Estimated instantaneous transition matrix under the null model ... 140  
7.12 Instantaneous-compounded one-year transition matrix approximation under the continuous null model .................................................. 141  
7.13 Continuous one-covariate stability-migration model results .......... 142  
7.14 Continuous one-covariate upgrade-downgrade model results .......... 143  
7.15 Top ten best continuous two-covariate model results ............... 144
Part I

A Study on Credit Cycle and Business Cycle
Chapter 1

Introduction

1.1 Motivation

The business cycle is a well-accepted economic fact. It describes the recurrent behavior of business conditions characterized by the occurrence order of several key macro-economic indicators; primarily, they are economic growth indicator GDP (gross domestic product), inflation indicators including PPI (producer price index) and CPI (consumer price index), employment indicators like non-farm payroll number and unemployment rate, etc.

Other than these macro-economic variables measuring the whole business condition of an economy, there is another group of variables which might be business condition related but particularly reflect the changes in economic credit conditions. One example is borrowers’ average default rate, measuring the proportion of borrowers default to overall borrowers with outstanding debt.

For the first part of our research, we try to answer following three questions:
1. Is there a recurrent phenomenon which we call credit cycle?

2. If the answer to the above question is positive, then whether or not credit cycle and business cycle can make up a whole economic cycle?

3. If the answer to the second question is positive, then what is the leading-lagging relation between credit cycle indicators and business cycle indicators in a whole economic cycle?

In this Part I study, we emphasize on exploring the relation between credit cycle and business cycle. In the following two chapters, we introduce the raw data of credit variables and business variables and their modifications; we elaborate how to use Complex Singular Value Decomposition (CSVD) method to detect the relative lead-lag relation among those indicators; we present and interpret CSVD results at the end of Part I.

1.2 Data description

All time series are obtained from FRED (Federal Reserve Economic Data) website \(^1\), except for the annual speculative-grade issuers’ default rate, obtained from the commercial rating agency S&P (Standard&Poor’s) rating report [13]. In this Part I study, the time horizon studied is from January 1987 to December 1999.

According to the primary economic meaning embodied in each time series, we classified all 17 time series into three categories: business variables, credit variables and credit stress variables. Specifically, business variables (GDP, interest rates, PPI, 

\(^1\)http://www.stls.frb.org
CPI, unemployment rate and payroll) reflect the stage of business cycle while credit variables (default rate, delinquency rate and charge-off rate) imply how the overall credit condition is. The complex correlation between business events and credit events make some time series hard to classify simply into business variable set or credit variable set. This is because not only business condition but also credit health have significant impact on their values. Therefore, we categorized them as credit stress variables, including the spread between Baa and Aaa bond yields, and household debt burden.

1.2.1 Business variables

- GDP – real gross domestic product
  Measured in chained 1996 dollars which reflecting real prices and buying power relative to year 1996.

- Interest rates
  - Short rates
    * R6M – yield of six-month treasury bill
    * GS01 – yield of one-year treasury bill
    R6M and GS01 are observed from the secondary market in which securities are traded after they are initially offered.
  - Long rates
    * GS10 – yield of ten-year treasury bond
GS30 – yield of 30-year treasury bond

GS10 and GS30 are zero-coupon bond rates.

- PPI – core producer price indexes (finished goods less food and energy)
- CPI – core consumer price indexes (finished goods less food and energy)
- UNEMP – unemployment rate
- PAYRL – number of employees non-farm payrolls

Additionally, GDP, PPI, CPI, UNEMP and PAYRL are all seasonally adjusted data. Seasonal adjustment removes the change that is due to normal seasonal activities, thus leaving an change only reflect economic trend and irregular movements.

1.2.2 Credit variables


  It is an observed default rate measure only on speculative-grade issuers, in that speculative-grade default occurrence is more sensitive to economic condition than investment-grade. Speculative-grade issuers are those debt issuers having deteriorating credit quality i.e. having high default probability.

- Delinquency loan rates
  - DRE – for real estate loans
  - DCI – for commercial and industrial loans
Delinquent loans are those past due thirty days or more and still accruing interest as well as those in nonaccrual status. They are measured as percentage of end-of-period loans.

- Charge-off rates
  - CRE – for real estate loans
  - CCI – for commercial and industrial loans

Charge-offs, which are the value of loans removed from the books and charged against loss reserves, are measured net of recoveries as a percentage of average loans and annualized.

1.2.3 Credit distressed variables

- Baa spread – yield difference between Moody’s Baa-rated bond and Aaa-rated bond

The reason to use Aaa as the benchmark interest rate for Baa instead of riskless Treasury yield is because of the discrepancies between corporate bond and Treasury bond in terms of liquidity and tax issues.

- Household debt-service burden – percentage of disposable personal income
  - CONS – consumer
  - MORG – mortgage
Disposable personal income is the amount of personal income an individual has after taxes and government fees, which can be spent on necessities, or non-essentials or be saved.

1.2.4 Data preparation

We notice that the observing time intervals are different. S&P overall default rate is reported annually; GDP, DRE, DCI, CRE, CCI, CONS and MORG are recorded quarterly; the other time series have monthly observations. Time consistency is needed to carry out this series analyses of these data. Since most of data have monthly interval, we adopt one month as a time unit for all the data. That is, to expand quarterly raw data to monthly data, we equalize their artificial monthly data to the corresponding quarterly observation. For annual default rate, we derived the monthly default rate by cubic spline approximation.

Raw data of GDP, PAYRL, CPI and PPI are some quantitative indexes which have no indication meaning of economy stage. In practice, the change rates of these time series relative to previous observation are used as business condition indicator. Therefore, we use the first difference of log of time series for this study.

Finally, we standardize all 17 time series by subtracting their sample mean and dividing them by their standard deviation. And the data matrix is denoted as $X_{n \times p}$, where contains $n = 1, 2, \ldots, 156$ monthly observations for $p = 17$ variables described as above.
Chapter 2

Complex Singular Value Decomposition Analysis

To detect the propagating relation between credit cycle and business cycle, we approximate the data matrix $X_{n \times p}$ with the Complex Singular Value Decomposition (CSVD) method.

To ensure no relevant variables are omitted, all the modified time series are included into the data matrix $X_{n \times p}$, which represents the $n = 156$ monthly observations during 1983 – 1999 on each of the $p = 17$ variables. Rows denote equally-distanced time points $t = 1, 2, \ldots, 156$ and columns $x_1, x_2, \ldots, x_p$ represent $p = 17$ different variables.

Because the resulting data matrix $X_{156 \times 17}$ is large, it is very difficult to analyze. Naturally, the Singular Value Decomposition method (SVD) can be used to reduce the original data set into a smaller number of significant components which are linear combinations of observable variables and account for the largest variations. There-
Therefore, the data matrix can be approximated by a form of SVD. However, real-value SVD is that it incorporates only variation information not the lead-lag information among variables in a complete cycle. Essentially, SVD theory is exactly same as Principal Component Analysis (PCA). Brillinger (1981) ([14] and Horel (1984) [23] show that an alternative to real-valued PCA is Complex Principal Component Analysis (CPCA), which allows efficient detection of cyclic features in all the time series. Correspondingly, we adopt SVD’s complex version as Complex Singular Value Decomposition (CSVD) in this study. Next, we explain how to transform real observations to complex values.

2.1 Complex value transformation

Denote the $t$ th monthly observation of the $j$ th variable as $x_{t,j}$, $t = 1, 2 \ldots, 156$; $j = 1, 2, \ldots, 17$. Because any reasonable periodic function can be rewritten as a linear combination of cosine functions, we suppose a periodic time series $x_{t,j}$ that has a smooth signal:

$$x_{t,j} = R_{t,j} \cos (2\pi f_j t + \phi_j); \quad (2.1)$$

where $R_{t,j}$ is the amplitude of the $j$th variable at time $t$; $\phi_j$ is the starting phase angle and $f_j$ is the frequency. Moreover, its complex form is:

$$x_{t,j} = R_{t,j} \cos (2\pi f_j t + \phi_j)$$

$$= \frac{1}{2} \left\{ (R_{t,j}e^{i(2\pi f_j t + \phi_j)} + R_{t,j}e^{-i(2\pi f_j t + \phi_j)} \right\} \quad (2.2)$$

$$= \frac{1}{2} R_{t,j} e^{i\phi_j} e^{i2\pi f_j t} + \frac{1}{2} R_{t,j} e^{-i\phi_j} e^{-i2\pi f_j t}$$

$$= x_{t,j}^c + x_{t,j}^c.$$
Thus, any real-valued $x_{t,j}$ is decomposed into two conjugate complex parts: $x_{t,j}^c$ and $\overline{x_{t,j}}$. Obviously, $x_{t,j}^c$ and $\overline{x_{t,j}}$ have a one-to-one realtionship. Here, we choose $x_{t,j}^c$ as the complex form of time series we needed. Again, by above (2.2),

$$x_{t,j}^c = \frac{1}{2} R_{t,j} e^{i \phi_j} e^{i 2\pi f_j t}. \quad (2.3)$$

Because amplitude $R_{t,j}$, frequency $f_j$ and phase $\phi_j$ are not observable, we can not derive this complex form directly from above equation (2.3). Next, we will explain how to transform the real-valued data into this positive frequency complex-valued time series by using the property of Discrete Fourier Transform (DFT) and Inverse Discrete Fourier Transform (IDFT). See Bloomfield (2000) [12] for details.

For $n$ observations equally spaced, the Fourier frequencies are

$$\omega_k = \frac{k}{n} \quad \text{for } k = 0, 1, \ldots, n - 1.$$

Let $\mathbf{y}$ be a time series vector with $n$ observations,

$$\mathbf{y} = (y_1, y_2, \ldots, y_n).$$

The DFT of $\mathbf{y}$ is $f(\omega_k)$, which is formulated as

$$f(\omega_k) = \frac{1}{n} \sum_{s=1}^{n} y_s e^{-i 2\pi \omega_k s}, \quad \text{for } k = 0, 1, \ldots, n - 1. \quad (2.4)$$

And the IDFT of $f(\omega_k)$ is, for $t = 1, 2, \ldots, n$,

$$\sum_{k=0}^{n-1} f(\omega_k) e^{i 2\pi \omega_k t} = \sum_{k=0}^{n-1} \frac{1}{n} \left( \sum_{s=1}^{n} y_s e^{-i 2\pi \omega_k s} \right) e^{i 2\pi \omega_k t},$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} \sum_{s=1}^{n} y_s e^{-i 2\pi \omega_k (s-t)},$$

$$= \frac{1}{n} \sum_{s=1}^{n} y_s \left( \sum_{k=0}^{n-1} e^{-i 2\pi \omega_k (s-t)} \right). \quad (2.5)$$
Because
\[
\sum_{k=0}^{n-1} e^{-i2\pi\omega_k(s-t)} = \begin{cases} 
0 & \text{if } s \neq t, \\
1 & \text{if } s = t.
\end{cases}
\] (2.6)

IDFT (2.5) becomes
\[
y_t = \frac{1}{n} \left( y_t \times n + \sum_{s \neq t} y_s \times 0 \right) = y_t.
\] (2.7)

In summary, by above DFT and IDFT relationship (2.4), (2.5) and (2.7),
\[
y = \text{IDFT} \circ \text{DFT} \circ y.
\]

Therefore, the \(t\)th observation of \(j\)th time series \(x_{t,j}\) becomes the summation of two conjugate complex terms. That is,
\[
x_{t,j} = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{s=1}^{n} x_{s,j} e^{i2\pi\omega_k(t-s)},
\]
\[
= \bar{x}_{t,j} + x_{t,j}^1 + x_{t,j}^2.
\]

Since the sample mean \(\bar{x}_{t,j} = 0\) from the data standardization, then
\[
x_{t,j} = x_{t,j}^1 + x_{t,j}^2,
\] (2.8)

where
\[
x_{t,j}^1 = \begin{cases} 
\frac{1}{n} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \sum_{s=1}^{n} x_{s,j} e^{i2\pi\omega_k(t-s)} & \text{if } n \text{ is odd}, \\
\frac{1}{n} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \sum_{s=1}^{n} x_{s,j} e^{i2\pi\omega_k(t-s)} + \frac{1}{2n} \sum_{s=1}^{n} x_{s,j} e^{i\pi(t-s)} & \text{if } n \text{ is even}
\end{cases}
\] (2.9)

and
\[
x_{t,j}^2 = \begin{cases} 
\frac{1}{n} \sum_{k=\lceil \frac{n}{2} \rceil+1}^{n-1} \sum_{s=1}^{n} x_{s,j} e^{i2\pi\omega_k(t-s)} & \text{if } n \text{ is odd}, \\
\frac{1}{n} \sum_{k=\lceil \frac{n}{2} \rceil+1}^{n-1} \sum_{s=1}^{n} x_{s,j} e^{i2\pi\omega_k(t-s)} + \frac{1}{2n} \sum_{s=1}^{n} x_{s,j} e^{i\pi(t-s)} & \text{if } n \text{ is even}
\end{cases}
\] (2.10)
Obviously, $x^1_{t,j}$ and $x^2_{t,j}$ are conjugate to each other. This is because when $n$ is odd,

\[
x^2_{t,j} = \frac{1}{n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} x_{s,j} e^{i 2\pi \omega_k (t-s)},
\]

\[
= \frac{1}{n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} x_{s,j} e^{i 2\pi (1-\omega_k)(t-s)},
\]

\[
= \frac{1}{n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} x_{s,j} e^{-i 2\pi \omega_k (t-s)},
\]

\[
= \overline{x^1_{t,j}},
\]

and when $n$ is even,

\[
x^2_{t,j} = \frac{1}{n} \sum_{k=\frac{n}{2}+1}^{n-1} \sum_{s=1}^{n} x_{s,j} e^{i 2\pi \omega_k (t-s)} + \frac{1}{2n} \sum_{s=1}^{n} x_{s,j} e^{i \pi (t-s)},
\]

\[
= \frac{1}{n} \sum_{k=\frac{n}{2}+1}^{n-1} \sum_{s=1}^{n} x_{s,j} e^{i 2\pi (1-\omega_k)(t-s)} + \frac{1}{2n} \sum_{s=1}^{n} x_{s,j} e^{i \pi (t-s)},
\]

\[
= \frac{1}{n} \sum_{k=\frac{n}{2}+1}^{n-1} \sum_{s=1}^{n} x_{s,j} e^{-i 2\pi \omega_k (t-s)} + \frac{1}{2n} \sum_{s=1}^{n} x_{s,j} e^{-i \pi (t-s)},
\]

\[
= \overline{x^1_{t,j}},
\]

So the two summation parts $x^1_{t,j}$ and $x^2_{t,j}$ of the above equation are conjugate to each other. And $x^1_{t,j}$ contains positive Fourier frequencies and $x^2_{t,j}$ has negative Fourier frequencies.

In fact, $x^1_{t,j}$ is the complex transformation of $x_{t,j}$ which we called $x^c_{t,j}$ in (2.3).
This is because

\[
x^1_{t,j} = \frac{1}{n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} x_{s,j} e^{i2\pi\omega_k(t-s)} ,
\]

for time series with form (2.1),

\[
= \frac{1}{2n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} \left( R_{t,j} e^{i(2\pi f_j s + \phi_j)} + R_{t,j} e^{-i(2\pi f_j s + \phi_j)} \right) e^{i2\pi\omega_k(t-s)},
\]

\[
= \frac{1}{2n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} R_{t,j} e^{i(2\pi f_j s + \phi_j + 2\pi\omega_k(t-s))} + \frac{1}{2n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} \sum_{s=1}^{n} R_{t,j} e^{-i(2\pi f_j s + \phi_j - 2\pi\omega_k(t-s))},
\]

\[
= \frac{1}{2n} R_{t,j} e^{i\phi_j} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} e^{i2\pi\omega_k t} \sum_{s=1}^{n} e^{i2\pi(f_j - \omega_k)s}
\]

\[
+ \frac{1}{2n} R_{t,j} e^{-i\phi_j} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} e^{i2\pi\omega_k t} \sum_{s=1}^{n} e^{-i2\pi(f_j - \omega_k)s}.
\]

(2.11)

If we assume there exists a \(k_0 \geq 0\) such that \(f_j = \omega_{k_0}\), then \(\omega_{k_0} - \omega_k = \omega_{k_0-k}\) and \(\omega_{k_0} + \omega_k = \omega_{k_0+k}\) for \(k > 0\) are still Fourier frequencies. For large \(n\), this assumption is reasonable. Thus, we have

\[
\sum_{s=1}^{n} e^{-i2\pi(f_j - \omega_k)s} = \sum_{s=1}^{n} e^{-i2\pi\omega_{k_0-k}s} = \begin{cases} n & \text{if } k = k_0; \\ 0 & \text{if } k \neq k_0; \end{cases}
\]

(2.12)

and

\[
\sum_{s=1}^{n} e^{i2\pi(f_j + \omega_k)s} = \sum_{s=1}^{n} e^{i2\pi\omega_{k_0+k}s} = 0, \quad \text{for } k_0 \geq 0 \text{ and } k > 0.
\]

(2.13)

Finally, \(x^1_{t,j}\) in (2.11) becomes,

\[
x^1_{t,j} = \frac{1}{2} R_{t,j} e^{i(2\pi f_j t + \phi_j)} + 0
\]

\[
= x^c_{t,j}.
\]
So, we know that
\[ x_{t,j}^1 = \frac{1}{n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} n \sum_{s=1}^{n} x_{s,j} e^{i2\pi \omega_k (t-s)} \]
is the complex transformation we need. In summary,
\[ x_{t,j}^1 = \frac{1}{n} \sum_{k=1}^{\left\lfloor \frac{n}{2} \right\rfloor} n \sum_{s=0}^{n-1} x_{s,j} e^{i2\pi \omega_k (t-s)} = \frac{1}{2} R_{t,j} e^{i\phi_j} e^{i2\pi f_j t} = x_{t,j}^c \]

When \( n \) is odd, we have similar discussion and we can conclude that
\[ x_{t,j}^1 = \frac{1}{n} \sum_{k=1}^{\frac{n-1}{2}} n \sum_{s=0}^{n-1} x_{s,j} e^{i2\pi \omega_k (t-s)} + \frac{1}{2n} \sum_{s=0}^{n-1} x_{s,j} e^{i\pi (t-s)} = \frac{1}{2} R_{t,j} e^{i\phi_j} e^{i2\pi f_j t} = x_{t,j}^c \]

Thus, the complex transformation on real data is accomplished by the IDFT with positive Fourier frequencies on DFT of each time series.

### 2.2 Complex singular value decomposition analysis

By CSVD, a full-ranked complex-valued data matrix \( X_{n \times p}^c, p \leq n \), can be decomposed as
\[
X_{n \times p}^c = U_{n \times p} D_{p \times p} V_{p \times p}^* \tag{2.14}
\]
where \( D_{p \times p} \) is a diagonal matrix with positive real-valued diagonals \( d_1, d_2, \ldots, d_p \), which are the singular values of matrix \( X_{n \times p}^c \). We assume these diagonals are in non-increasing order. The complex column vectors \( (u_1, u_2, \ldots, u_p) \) of \( U \) and the complex column vectors \( (v_1, v_2, \ldots, v_p) \) of \( V \) are orthogonal and unitary:
\[
U^* U = V^* V = I_{p \times p} \tag{2.15}
\]
where \( \ast \) denotes conjugate transpose. Hence, the columns of \( U \) form \( p \) orthogonal basis vectors for the columns of \( X^c_{n \times p} \) in \( n \)-dimensional space and the columns of \( V \) form an orthogonal basis for the rows of \( X^c_{n \times p} \) in \( p \)-dimensional space. Vector \( u_i \) is called the \( i \)th score and vector \( v_i \) is called the \( i \)th loading, \( i = 1, 2, \ldots, p \), of \( X^c_{n \times p} \).

A data matrix can be approximated by replacing original \( p \) variables by fewer number of variables, say, \( r < p \), according to the structure of above CSVD (2.14). That is, CSVD approximates \( X^c_{n \times p} \) as

\[
X^c_{n \times p} \simeq X^c_{n \times p} = U_{n \times r}D_{r \times r}V^*_p = \sum_{i=1}^{r} d_i u_i v_i^*.
\]

Score vectors \( u_i \) and loading vectors \( v_i \) can be derived from the CPCs (complex principal component) of matrices \( X^c \) and \( X^c \), respectively.

\[
X^c = V_{p \times p} D_{p \times p} U^*_n
\]

CPCA can be used to determine \( r \) new variables i.e. CPCs which are the orthogonal linear combinations of original variables and which also account for most of variation in the original variables. Data matrix \( X^c \) can be considered as the \( n \) observations of \( p \) original variables and its transposed complex conjugate matrix \( X^c \) can be considered as the \( p \) observations of \( n \) different time variables. Both sets of original variables are of interest for we want to understand both time structure and variable structure of the whole cycle.

Because of the normalization on time series \( x_j \), complex transformed matrix \( x^c_j \) has zero sample mean and roughly \( \frac{1}{2} \) sample variance which is constant for every \( j \). So the \( X^c X^c \) is proportional to its sample correlation matrix for column variables in \( X^c \) and \( X^c X^c \) is proportional to its sample correlation matrix for row variables in \( X^c \).
From CSVD (2.14), we derived the CSVDs of matrices $X^c X^c$ and $X^c X^{c*}$ as

$$X^c X^c = VDU^*UDV^*$$

by (2.15),

$$= VD^2 V^*;$$

similarly,

$$X^c X^{c*} = UD^2 U^*.$$

For a positive definite symmetric matrix, CSVD is equivalent to the spectral decomposition. Hence, the loading vectors in $V$ are the eigenvectors of matrix $X^c X^c$ and the score vectors in $U$ are the eigenvectors of matrix $X^c X^{c*}$. In fact, matrices $X^c X^c$ and $X^c X^{c*}$ are the sample correlation matrices for $p$ times series and $n$ time points, respectively. Because of the Hermitian property of these two matrices, the eigenvalues are guaranteed to be positive. Moreover, they have same eigenvalues which are the squares of the singular values of $X^c$.

The principal components of matrix $X^c$ are $X^c V = UDV^* V = UD$; and the principal components of matrix $X^{c*}$ are $X^{c*} U = VDU^* U = VD$. By the orthogonality in matrices $U, V$, these PCs are orthogonal to each other. Specifically, for matrix $X^c$, the vector $d_i u_i \sim$ contains $n$ observations of the $i$th new variable which is the combination of $p$ original variables by linear operation $v_i \sim$ and this new variable counts the $i$th most correlation in original $p$ variables. Similarly, from the time variable point of view, the vector $d_i v_i \sim$ contains $p$ observations of the $i$th new variable
which is the combination of \( n \) original time variables by linear operation \( \mathbf{u}_i \) and this new variable counts the \( i \)th most correlation in those original \( n \) variables. And the proportion of correlation this \( i \)th new variable counts for is

\[
p_i = \frac{d_i^2}{\sum_{1}^{p} d_i^2}
\]  

(2.18)

If the proportion of correlation explained by the first CPC is high, we adopted \( r = 1 \) in CSVD approximation (2.16). This is true for our study and will be shown in next result subsection. Consequently, data matrix \( \mathbf{X}^c \) is approximated by both business/credit variable labels in \( \mathbf{v}_1 \) and the time variable labels in \( \mathbf{u}_1 \). That is, for \( r = 1 \),

\[
\hat{\mathbf{X}}_{n \times p}^c = d_1 \mathbf{u}_1 \mathbf{v}_1^*; \quad (2.19)
\]

where \( \mathbf{u}_1 = (u_{11}, u_{12}, \ldots, u_{1n})' \), and \( \mathbf{v}_1 = (v_{11}, v_{12}, \ldots, v_{1p})' \).

Because \( u_{1,t} \) is the \( t \)th complex observation in the \( n \)-observation vector \( \mathbf{u}_1 \) of the first CPC and \( v_{1,j} \) is the \( j \)th complex component of first CPC linear operator \( \mathbf{v}_1 \) transforming the original \( p \) variables to this unobservable first CPC. We express the complex score vector and loading vector as

\[
u_{1,t} = a_t e^{i2\pi F(t)} ,
\]

(2.20)

and

\[
v_{1,j} = b_j e^{i2\pi \phi_j}.
\]

(2.21)

For \( u_{1,t} \), \( a_t \) indicates the cyclic strength of all variables at time \( t \); period \( \frac{1}{F(t) - F(t-1)} \) can tell how many months a cycle takes. Meanwhile, for \( v_{1,j} \), \( b_j \) may be thought of as the loading contribution measurement of the original variable \( j \) to this CPC. So
we can decide the significance of this variable in cyclic analysis by the value of $b_j$. The phase angle contribution of variable $j$ is $\phi_j$. In fact, the relative phase difference between two different variables allows us to identify the their lead-lag relationship.

If $a_t$ is nontrivial at any time $t$, we can conclude that business cycle and credit cycle share same frequency and they can make up a complete economic cycle. Then, we are interested in the propagating relation among these business and credit variables. The complex form of the $j$th time series is

$$x^{c}_{t,j} \cong \frac{1}{2} R_{t,j} e^{i \phi_j} e^{i 2\pi F(t)}, \quad (2.22)$$

where $F(t)$ is the total phase change from time 0 to $t$, and $\frac{\partial F(t)}{\partial t} = f(t)$ is the instantaneous frequency at time $t$.

### 2.3 First order complex singular value decomposition result

In this subsection, we show some results of first order CSVD on all 17 time series and we explore the implications from score vector $u_1$ and loading vector $v_1$. The first CPC can explain as high as 63% of correlation among these 17 complex economic time series, which confirms that first order CSVD can give us a good approximation of a complex data matrix.

The appearance of the main economic trend is investigated first in the time domain. A whole economy includes both business conditions and credit conditions.

1. **First score**: $u_{1,t} = a_t e^{2\pi F(t)}, \quad t = 1, 2, \ldots, 156.$
Figure 2.1 shows the amplitude time series of the first score: \( \{a_1, a_2, \ldots, a_{156}\} \). Other than the two ending jumps (resulting from computing a DFT on a time series with discontinuity between the first and last real-valued observations), large amplitudes between 1990 and 1995 indicate that the whole economy’s cyclic trend is strong during that time period. The small amplitude since 1995 implies that economic recurrent appearance has been weaker since then.

Figure 2.2 is the graph of first score’s phase time series \( \{F(t), t = 1, 2, \ldots, 156\} \). The linear appearance between time \( t \) and phases \( F(t) \) is an existence evidence of cyclic trend of the whole economy, including both business condition and credit conditions. However, slopes look different for three periods: 1987-1991, 1991-1994, in the middle of 1995-present.

The monthly frequencies are calculated as from subtracting phases by their preceding monthly phase, i.e. \( \{f(t) = F(t) - F(t-1)\} \). The frequency time series during 1987–2000 is showed in Figure 2.3. Again, we notice that frequencies are more stable within these three periods and are more different between these periods. In summary, the change in economy with time is recurrent but has different periods for different cycles.

The periodicity of the whole economy over this period is calculated by \( \frac{1}{f(t)} \) and is visualized in Figure 2.4. Interestingly, there is some economic disturbance between late of 1994 and middle of 1995. And the period of whole economic cycle is larger since then.

Next, we look into the linear relationship between first CPC and all 17 macro-economic variables and credit variables.
2. First loading: \( v_{1,j} = b_j e^{i\phi_j}, \quad j = 1, 2, \ldots, 17. \)

Figure 2.5 is the first loading graph, which is a visual representation of the complex linear relationship between the original 17 variables and the unobservable first CPC. First CPC is a linear combination of these 17 variables and \( v_1 \) is the coefficient vector.

First, the amplitude of each normalized variable coefficient measures the weight of its linear contribution to the first CPC. The greater the amplitude, the more important the variable to detect the first CPC and to account for the economic cycle. Obviously, standardized inflation indicators dIPPI and dCPI are not as important as other standardized variables in investigating the cyclic trend of the first CPC.

Secondly, we notice that variables carrying similar economic information tend to be close to each other. Economy health indicators dGDP and dPAYRL stay close to the same line; short interest rates TBM06 and GS01 are almost identical in both phase and amplitude; similarly are the long rates GS10 and GS30; and all the credit variables (DEFR, DRE, CRE, DCI, CCI) cluster together to represent the credit cycle. Therefore, we roughly classified all 17 variables into nine clusters: productivity variable dGDP and the number of new jobs variable dPAYRL make up the economy growth cluster; R6M and GS01 belong to short rate cluster; GS10, GS30 and MORG are included in the long rate cluster; the corporate credit health cluster contains DCI and CCI; the consumer credit health cluster on real estate includes DRE and CRE; CONS, DEFR and UNEMP are three single-variable clusters.

Because dIPPI and dCPI have smaller amplitudes and they are in line with the long rate cluster, they can be either neglected or included in the long rate cluster. The relative phase positions of different clusters are very useful in interpreting their
lead-lag occurrence relationship during an economic cycle.

In a complete economic cycle, let us start with the change in the economic growth variable $\text{dlGDP}$. Then the order of changes in these clusters would be found by moving counterclockwise around Figure 2.5:

$\text{dlGDP}, \text{dlPAYRL} \rightarrow \text{CONS} \rightarrow \text{R6M}, \text{GS01} \rightarrow \text{SPREAD} \rightarrow \text{GS10, MORG, GS30} \rightarrow \text{DEFR} \rightarrow \text{DCI, CCI} \rightarrow \text{DRE, CRE} \rightarrow \text{UNEMP}$.

From the above order in Figure 2.5, short interest rates are more sensitive to the stage of economy than are long interest rates and SPREAD has intermediate importance; the consumer debt payment proportion variable CONS precedes short rates while the mortgage debt payment proportion MORG behaves similar to long interest rates; DEFR leads the other credit variables; the corporate loan health variables DCI and CCI change before the consumer real estate loan creditworthiness DRE and CRE.

If an economic cycle starts with the change in UNEMP, then all the business variables lead the credit variables. From the preceding study of the first score vector, we know that the economy as a whole is recurrent. That is, all the 17 business and credit variables compose a strong cyclic trend with time. This result makes it possible to model credit cycles by the well-known business cycle indicators. The related discussion is left for future work.

Figure 2.6 visually represents the order of occurrence of economic variables within one complete economic cycle.

The cluster of $\text{dlGDP}$ and $\text{dlPAYRL}$ is a measurement of economic growth. One complete economic cycle can be broken into four stages: economic expansion, economic slowdown, economic recession, and economic recovery.
From Figure 2.6, when the economy is expanding, economy productivity and new non-agriculture jobs grow at an increasing speed; unemployment rate keeps dropping at this stage; short rates start to increase from their bottoms; long rates stop subsiding at the end of this expansion; a larger proportion of consumers’ income is used to pay their consumption debt; at the same time, credit condition is healthy: there is low proportion of speculative-rated issuers defaulting; less and less corporate loans and real estate loans become delinquent or get charged-off.

When the economy is in the stage of slowdown, both short rates and long remain increasing; employment rates start to get higher; consumers still have an increasing proportion of consumption debt payment and mortgage payment to their disposable income; the credit condition starts deteriorating: more and more credit risk is indicated by an increase in SPREAD; more and more corporations and individuals fail to pay a payment on their loans; more and more speculative-graded debt issuers default at the same time.

In the economy recession, productivity and non-agricultural employment continue growing negatively; the unemployment rate remain increasing; interest rates begin declining in the middle of this stage with short rates preceding to long rates; consumers pay less debt relative to their disposable income; credit situation still is worsening but starts improving at the end of this recession: SPREAD and default rate turn to decrease and less proportion of loans is charged off or become delinquent.

In economic recovery, GDP has better growth rates even though still negative. At the same time, both business condition and credit condition are improving: interest rates are decreasing; unemployment rate begins decreasing; SPREAD becomes nar-
rower; default rate decreases; and more corporations and individuals are capable to pay off their required loan payments.

2.4 Summary

In Part I study, we use Complex Singular Value Decomposition (CSVD) analysis to investigate the credit condition trend in U.S. economy and its leading-lagging relation with business condition. We find that the whole economy represented by business condition along with credit condition is periodically varying with time. But the recurrent frequencies are different for different time intervals during 1988 – 1999. Roughly speaking, U.S. economy has greater cyclic periodicity after 1995 than before 1995.

The well fit-in of credit time series into the whole economic cycle indicates that the change in credit condition is recurrent. Thus, the existence of credit cycle is proved. Moreover, the clustering property of all credit variables in a cycle in Figure 2.6 identify the leading-lagging relation between credit cycle and business cycle. If we start an economy cycle with business variable UNEMP, then all the business variables are preceding to credit variables. Therefore, we may be able to model and prediction credit cycle with business cycle indicators. We will leave this problem to a future work.
Figure 2.1: Amplitude of first score vector
Figure 2.2: Phase time series of the first CPC
Figure 2.3: Frequency time series of the first CPC
Figure 2.4: Period time series of the first CPC
Figure 2.5: Loadings of the first CPC
Figure 2.6: First CPC representation of one economy cycle
Part II

A Generalized Linear Model of Business Condition on Credit Risk

Drift Process
Chapter 3

Introduction

In the Part II of this dissertation, we study the probability that borrowers will default on their financial commitments within a given time period. This chapter describes the general definition of credit risk and raises the question of how to incorporate business impact into credit risk modeling. Section 3.1 introduces the concept of credit risk and motivates the need of proper credit risk measure. Section 3.2 reviews the advantages and disadvantages of different types of approaches to modeling credit risks. Section 3.3 presents the outline of our Part II research.

3.1 Background

1. What is credit risk

Credit risk is generated at the moment that a debt contract is agreed between a borrower and a lender; it will continue existing until the borrower pays off all the financial obligations specified in the contract. Credit risk is the chance that a
borrower may fail to fulfill its commitment to the lender in the remaining life of their contract. In such an event, the lender suffers a financial loss. This kind of failure to make agreed payment is called default. A borrower is the counterparty to a lender, and vice versa.

During the lifetime of a contract, a lender is always exposed to its counterparty’s potential default risk, i.e. credit risk. This is because a borrower’s default probability is not zero at any given time if the contract is alive. Consequently, a positive return or profit on a contract becomes a compromise between the needs from both counterparties. For borrowers, return is used as an incentive to attract capital investment; while for lenders, it is serving as an investment and as a financial compensation to bear borrowers’ default potential. For this reason, a borrower can also be called a debt-issuer; and a lender as an investor. In practice, return is usually quantified by yield to maturity. Yield (to maturity) is the discount rate that equates the present value of interest payments and redemption value with the price of the financial contract. We use a simple example to illustrate this definition: a bond is issued at time zero with price $b$ and is going to pay back $F$ in $t$ unit time. The yield $r \geq 0$ is calculated as,

\[ b = \frac{F}{(1+r)^t}. \]

Note that a contract’s yield varies inversely with its present price.

Different issuers have different amounts of credit risk: issuers with low credit risk (small default possibility) are considered to have good creditworthiness or high credit quality; those with high credit risk (great default possibility) are considered to have bad creditworthiness or low credit quality. Serving as a compensation for investors to
bear counterparty’s credit risk, yield should reflect the degree of issuer’s credit risk. That is, the higher the default possibility, the higher the yield on the contract, and thus the lower the contract price. Therefore, pricing any credit-risky financial contract should take into account an accurate measurement of issuer’s creditworthiness.

2. What are the existing credit risk measurements

In practice, there are two observable credit risk measurements: credit rating and credit risk spread.

Credit risk spread is the yield difference between a credit-risk-free instrument and a credit-risky one with same time to maturity. Usually, the risk-free credit instruments refer to Treasury bonds issued by U.S. government because of its never-default credit history. So it is well believed that the yield of risk-free bond reflects and measures only market risk. A corporate issued bond is an example of credit risky instrument. The extra credit risk implied in a credit-risky instrument such as a corporate bond results in its higher yield than the credit-risk-free’s. That is, a spread is the capital compensation for investors to bear the extra credit risk additional to market risk. Obviously, credit spread is narrower for instrument from high credit-worthy counterparties and is broader for instrument from low credit-worthy counterparties.

A spread is a quantitative measure of market perspective of a counterparty’s default. It can be thought of a measure of credit risk out of the market perspective. If the market of a credit-related instrument is complete, i.e. , then the mark-to-market price of that instrument implies how market participants project its credit risk at that point. However, historically severe illiquidity or nontradability of most credit-sensitive products prevents spread observations from reflecting and measuring credit
risk. This could be the reason why most current commercial credit risk modeling tools (CSFP’s CreditRisk+ [3], JP Morgan’s CreditMetrics [2] and McKinsey & Co’s ProtfolioView) are relying on the other risk classification system – credit ratings.

A credit rating system broadly categorizes issuers according to their historical credit-worthiness. This kind of research is conducted either publicly by credit rating agencies or privately by individual financial institutions. Currently, there are several U.S. major rating agencies: Standard and Poor’s (S&P), Moody’s Investors Service, Fitch IBCA, and Duff and Phelps Credit Rating Co.

A credit rating system provides a categorical (nonquantitative) and subjective credit risk measure. The underlying rational is that all the issuers with same rating should have a similar amount of credit risk.

Historically, there has been a strong and consistent correlation between credit spread and credit rating. Caouette et al (1998) [16] shows that the five-year default rates are consistently higher for lower ratings in Moody’s; the spreads over 30-year Treasury bonds for the past 20 years for higher-rated bonds have been always less than those for lower-rated bonds. Their studies indicate that investors have applied the same credit rating information on pricing credit-sensitive bonds. The authors conclude rating agencies succeed in giving investors a relatively reliable guide to credit risk.

At any time, an issuer’s creditworthiness might either improve or deteriorate. A change in an issuer’s credit quality will lead rating agencies to adjust their perspective of this issuer’s future default likelihood and maybe to assign it a new credit rating. Since potential default probability is crucial to pricing credit instruments, the market
price of this issuer’s credit instruments will change according to the direction and distance of its credit rating migration.

In this Part II credit rating migration research, we employed S&P’s [13] issuer annual credit rating transition observations from 1980 to 1999.

“A Standard & Poor’s Issuer Credit Rating is a current opinion of an obligor’s overall financial capacity (its credit-worthiness) to pay its financial obligations. This opinion focuses on the obligor’s capacity and willingness to meet its financial commitments as they come due. It does not apply to any specific financial obligation, as it does not take into account the nature of and provisions of the obligation, its standing in bankruptcy or liquidation, statutory preferences, or the legality and enforceability of the obligation. In addition, it does not take into account of credit-worthiness of the guarantors, insurers, or other forms of credit enhancement on the obligation.”

Our S&P’s rating data include observed 19 one-year transition matrices for each year during 1980-1999 and the number of issuers rated at each rating at the beginning of each year.

An observed one-year transition matrix is made up by the transition frequencies from every initial rating at the year beginning to any final ratings observed at the year ending. There are seven observed initial ratings, if ranked from best credit rating to worst, which are: AAA, AA, A, BBB, BB, B, CCC. And nine observed year-end ratings: AAA, AA, A, BBB, BB, B, CCC, D and N.R.. D is default rating and N.R. is the state of not being rated. For any year, the rating $i$ to rating $j$ transition rate is calculated as the number of issuers with rating $i$ at year-beginning and rating $j$ at year-end divided by the total number of issuers with rating $i$ at year-beginning. So an observed one-year rating transition matrix is a $7 \times 9$ rate matrix with seven unitary row summations.
In fact, for any time period, the transition matrix can be similarly composed of the transition frequencies from seven initial ratings at the beginning of that time period to nine final ratings at the end of that time period. A transition matrix is a set of discrete observations, which records issuers’ rating only at two time points - the beginning and the ending. Any intermediate transition activity will be left out from a transition matrix. For example, an year-beginning $i$-rated issuer is firstly changed to a different rating $k$ and then is rerated to rating $j$, $j \neq k$. Its rating keeps at $j$ till year-end. In one-year transition matrix, this issuer is counted as one with transition from $i$ to $j$ and there is no any information on the path of this rating transition. We will further discuss the conflict between rating migration continuity and transition matrix discreteness property in Chapter 3.

3. Why modeling credit risk becomes more and more important

In recent years, with the increasing transaction of high-yield loan and the dramatic innovations of credit-related financial products such as credit derivatives, credit market has been growing rapidly. Therefore, how to precisely measure and appropriately manage credit risk gains more and more attention from both financial practitioners and regulators.

The main concern of financial regulators is to ensure that a financial institution’s capital reflects the risks that it is bearing. To avoid financial market crisis and to maintain its stability, BIS (Bank for International Settlements) established a regulatory model in 1988 Basle Accord. This regulation stated that all private-sector loans were subjected to an eight percent capital requirement. The major problem of this regulatory model is its neglect of two facts: different counterparties might have dif-

ferent amounts of credit risks; credit risks from multi-counterparties are correlated. Thus this percentage uniformity of capital requirement resulted in the inefficient usage of capital and ineffective control of the credit risk in financial market. Specifically, the eight percent capital requirement is excessive for counterparties with good creditworthiness; while it is far less than enough to prevent a financial crisis from happening for low credit quality counterparties.

In response to these dissatisfaction and ineffectiveness mentioned above, in the BIS 1996 Amendment to the Basel Accord, regulators allowed certain large banks to employ their own internal models to calculate capital for their risk exposure. This important change motivated banks to seek sound risk models which not only can maximize the interest of banks but also can better reflect the risk they are bearing.

In practice, two important factors need to be taken into account for the task of modeling and predicting credit risks. The neglect of any one factor can lead to substantial mismeasurement of credit risks.

One factor arises from the intrinsic relation between market risk and credit risk. As we demonstrated in Part I, credit variables and business variables as a whole can present a complete cycle. So any one variable can be a leading indicator to other variables if we view the whole cycle starting with the change in this variable. We also showed that credit index could be modeled by the other macro-economic business index, which agreed the results found by Altman and Kao(1992) [8], Carty (1997) [17]: business condition has impact on issuers’ credit risk. For example, a firm borrowed money from a bank and promised to pay interest periodically until it pays back all its debt at the time they agreed. When the state of economy is good, this
firm is more likely to be profitable and hence the default risk is low. At the same time, the bank is bearing less credit risk. In the other case, under a severe market condition, this same firm might be much less profitable and even not be able to meet its financial commitment. Then the credit risk tends to become higher than before.

The other factor is the measure of correlation between credit risks from different issuers. Instead at the level of one instrument, the risk is practically measured at the level of a portfolio - a pool of credit risky instruments. A portfolio’s credit risk measurement incorporates not only the risk consideration of each individual instrument’s marginal credit risk but also the risk correlation among them. Duffee (1996) [21] tests how dangerous is the neglect of the correlation factor in most credit modeling methodologies to risk measurement accuracy.

3.2 Literature review on modeling credit risk

With the explosive growth of credit market, researchers have managed to model and predict credit risks from different directions. Based on the characteristics of input variables used in the default probability models, there are primarily four types of credit risk modeling methodologies.

Type one is credit-scoring approach. In general, key accounting ratios are utilized to produce a credit risk score model or a measure of default probability. The model connecting observed account ratios and default risk can be either one or some combined usage of these statistical models: the linear probability model, the logit model, the probit model, and the discriminant analysis model. For detailed examples, see Atlman’s five variable model (1968) [5], a logit and discriminant analysis model in

The second type of credit risk model is called Contingent Claim Analysis (CCA) by Jones et al (1984) [29]. Basically, CCA is a generalization or extension of Black-Scholes’ option pricing model pioneered by Black and Scholes (1973) [11]. A borrower’s liabilities are viewed as contingent claims issued against underlying asset value which is assumed to follow a stochastic differential process with known expected rate of asset return and variance of return. Merton (1974) [35] defines a default event occurrence when a borrower’s asset value falls below its total liability. The model is well known as Black-Scholes-Merton model. Since then, similar CCA models are proposed by Black and Cox (1976) [10], Santomero and Vinso (1977) [38], Scott (1981) [39], Chance (1990) [18] and Shimko et al (1993) [40]. Hull and White (1995) [24], redefine default process as the borrower’s asset value hit some specified boundary. A current commercial software employing CCA modeling is KMV (1993) [1].

CCA credit risk models have the advantage of theoretical simplicity. But they are very difficult to implement in practice. The difficulties arise from the estimate of constant variation on asset return; from the structure complexity of different liabilities; from lack of ability to price credit derivatives whose values depend on credit ratings.
For more comments and criticism, see Jarrow (1997) [27] and Duffee (1996) [21].

The third type of model is developed to derive the implied default probabilities from the term structure of yield spread between default risky debt and risk free Treasury bond. A yield spread term structure relates time to maturity and corresponding yield spread. The spread models calculate the market projected default probability at any given time in the future. Term structure models can be found in Jonkhart (1979) [30], and Iben and Litterman (1989) [25]. One major concern about these model is the impractical assumption on the availability of risky yield observation because most credit instruments are highly illiquid or nontradable. As a result, applying spread models has been hampered both by the lack of market yield observations of credit risky products and by the observation inaccuracy from non-open market.

The fourth type is called rating migration analysis, which models credit rating migration probabilities within a given time by assuming rating transition follows a discrete Markovian process. The finite state space of the Markov process is made up of all the possible credit ratings. Rating transitions are assumed to have this property: given the history of an issuer’s credit ratings, the distribution of its future rating only depends on the information of its current rating. Such rating migration models are proposed by Jarrow and Turnbull (1997) [27], JP Morgan & Co.’s Credit Metrics (1997) [2], Nickell et al (2000) [36].

JP Morgan&Co.’s commercial credit risk modeling tool, CreditMetrics (1997), assumes stationary Markovian transition probabilities and estimates the one-year transition matrix by approximating different ratings’ multi-year (from one-year to 15-year) average default rates in a least square sense. Under the Markovian transition
assumption, multi-year transition can be expressed as the product of square single-year matrices for the years within that multi-year period. To measure the joint credit risk of two issuers, CreditMetrics borrows correlation of asset returns as the correlation between their credit risks. Credit Metrics shows that Markov chain is a reasonable modeling tool in that it fits the observed default rates well.

Two problems in CreditMetrics’ approach are noticed. Firstly, Jarrow and Turnbull (2000) [28] and Crouhy et al (2000) [20] criticize CreditMetrics for ignoring the impact of market condition on credit rating transitions. Carty (1997) [17] documents the fact that rating transition matrices vary according to the stage of business cycle. Our result of Part I credit cycle research is evidence for the strong cyclic correlation between business cycle and credit cycle. Therefore, the stationarity assumption of Markovian transition process is inconsistent with the empirical evidence and the current credit practice. Secondly, Jarrow and Turnbull (2000) [28] examine the accuracy of using correlation between two issuers’ asset return to represent the correlation between their correlation of movement in credit qualities and conclude the invalidity of this equal correlation assumption.

Different from CreditMetrics, another Markovian rating migration model built by Jarrow and Turnbull (1997) [27] incorporates market impact into modeling transition probabilities. They impose some market condition proportional adjustment $\pi_i(t)$ on instantaneous stationary (i.e. market independent) Markovian transition probability $q_{i,j}$, say, for transition $p_{i,j}(t, t + 1)$ from rating $i$ to $j$ in year $t$. the expression is,

$$p_{i,j}(t, t + 1) = \pi_i(t)q_{i,j};$$
with constraints:

\[ \sum_j p_{i,j}(t, t + 1) = 1 \text{ for all } t, i. \]

The market adjustment \( \pi_i(t) \) is deterministic and independent of final rating \( j \). Jarrow and Turnbull (1997) [27] notice that this independence of \( \pi_i(t) \) and transition direction may not be true in practice. This is because the risk premiums for upgrades and downgrades from the same rating are equivalent under that independence. Another assumption they make is that the correlation between two issuers’s rating transitions is driven by business condition.

Conditioning on the business state, issuers’ industrial belongings, and their domicile property, Nickell et al (2000) [36] assume rating transitions follow independent multinomial distributions for a single period. They find a significant difference between transition matrices for different stages of the business cycle: trough, peak and normal. They further model the independent transition probabilities conditional on business condition with ordered probit analysis and use maximum likelihood estimation to estimate business covariate coefficients.

Another commercial software for credit risk measurement is CreditRisk+ by Credit Suisse Financial Products (1997) [3]. CreditRisk+ assumes that, conditionally on the business cycle, transitions for different issuers are independent. Carty (1997) [17] empirically examines the existence of credit quality correlation and verifies that the strength of correlation is dependent on market condition.

In summary, for rating migration approach, the considerations of rating migration’s market dependence and correlation between two issuers’ rating transitions need to be taken into account.
Besides the above four major types of credit risk models, researchers also employed other various methods to accomplish the task of credit risk modeling. Among them, there are neural network analysis (Coats and Fant (1993) [19], Altman et al (1995) [9]), mortality approach (Altman (1989) [7]), default jump process (Duffie and Huang (1996) [22]). Because of the concern of data accessibility, in this study, we adopt the rating transition method to model credit risk drift and its dependence on business condition.
3.3 Research outline for Part II

In this study, our interest is to build up a sound model to measure debt issuers’ credit risks by incorporating the impact from business condition. As academic researchers, we have difficulty in accessing all the issuers’ accounting data and yield spreads of their credit instruments. Therefore, we adopted credit rating migration methodology to derive credit risk measurement. We elaborate our approach in the succeeding chapters.

Chapter 4 describes the rating data from S&P and modification of 12 macro-economic time series during the time period 1981-1999. In the second part of Chapter 4, we state four fundamental assumptions about rating transitions which are critical to modeling and estimating in both discrete and continuous time frameworks.

The time horizon for transitions modeled in Chapter 5 is discrete as one year period. Under the assumptions made in Chapter 4, we present a transition probability model varying with business condition which is represented by one or two of 12 macro-economic variables and estimate the model by maximum likelihood estimation.

We propose an instantaneous transition probability model in Chapter 6, which has a similar form of business impact to that in the discrete model. For both Chapter 5 and Chapter 6, we also discuss the issues of the $\chi^2$ goodness-of-fit test, over-dispersion consideration and statistical hypothesis tests. These tests would allow us to detect the significance of business impact on rating transition probabilities and identify the most important macro-economic variables. The estimation and test results are revealed in Chapter 7.

Finally, we summarize the motivation, methodology and statistical result in Chap-
ter 8. Some future work is also brought up for discussion, and the contributions of this study to financial practice are emphasized.
Chapter 4

Data and Assumptions

With the acknowledgement of importance of credit transition distribution in measuring credit risk, we reviewed several methods of modeling a rating transition matrix and compared the merits and drawbacks in the last Chapter. Here, we propose a new method which not only carries on the benefits of existing models but also overcomes some disadvantages of them.

This chapter has two parts. The first part contains the data description and modification for dependent and independent variables. The second part contains discussion of two vital assumptions on rating transition distribution that are the foundation of this study.

4.1 Data description

From the statistical point of view, data were grouped into response variable observations and independent variable observations. We used the S&P one-year rat-
ing transition matrices during 1981-1999 as the multivariate response variable and 12 macro-economic variables, from Federal Reserve Economic Data, as independent variables.

### 4.1.1 Credit rating transition matrix

Every year S&P publishes a rating performance report of its rated long-term corporate and sovereign issuers. In the present study, we employed the data from the 1999 reports in which S&P “analyzed the rating histories of 8,693 issuers of long-term ratings from Jan. 1, 1981, to Dec. 31, 1999” [13]. Table 4.1 is an example of one observed one-year transition matrix which has all the transition rates in year 1999.

<table>
<thead>
<tr>
<th>Initial rating</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
<th>D</th>
<th>N.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>88.14</td>
<td>3.95</td>
<td>0.00</td>
<td>0.56</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>7.34</td>
</tr>
<tr>
<td>AA</td>
<td>0.16</td>
<td>88.22</td>
<td>6.21</td>
<td>0.32</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.16</td>
<td>4.94</td>
</tr>
<tr>
<td>A</td>
<td>0.00</td>
<td>2.24</td>
<td>86.50</td>
<td>5.30</td>
<td>0.08</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
<td>5.72</td>
</tr>
<tr>
<td>BBB</td>
<td>0.00</td>
<td>0.38</td>
<td>3.81</td>
<td>85.61</td>
<td>3.53</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
<td>6.48</td>
</tr>
<tr>
<td>BB</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>2.85</td>
<td>79.82</td>
<td>6.47</td>
<td>0.52</td>
<td>1.03</td>
<td>9.18</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>0.34</td>
<td>1.72</td>
<td>78.35</td>
<td>3.55</td>
<td>7.33</td>
<td>8.48</td>
</tr>
<tr>
<td>CCC</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.53</td>
<td>0.00</td>
<td>2.53</td>
<td>56.96</td>
<td>0.38</td>
<td>7.59</td>
</tr>
</tbody>
</table>

S&P categorizes the credit worthiness of long-term issuers with ratings for their implied senior unsecured long-term debt. When an issuer has no senior unsecured debt, its rating is inferred from its other issues. This classification is based on an issuer’s capacity and willingness to meet its financial obligation. An issuer’s rating is derived from S&P’s professional evaluation on its future default probability. Canter
and Packer (1994) [15] and Caouette et al. (1998) [16] prove the relative reliability of this rating system based on historical transition observations.

In S&P issuers’ rating system, there are nine ratings: AAA, AA, A, BBB, BB, B, CCC, D and N.R.. D is default rating and N.R. is the state of not being rated. The order of other seven ratings from AAA to CCC indicate issuers’ credit qualities from best to worst. For example, at one time point, a CCC-rated issuer is considered to be more likely to default on its financial commitment than an AAA-rated one. The transition data in S&P’s 1999 report contain the transition observations from initial rating set \{AAA, AA, A, BBB, BB, B, and CCC\} to final rating set \{AAA, AA, A, BBB, BB, B, CCC, D and N.R.\}.

According to the inherent characteristics of debt issued by different rated issuers, initial ratings were roughly split into two categories: speculative-grade (BB, B, CCC) and investment-grade (AAA, AA, A, BBB). The yield spread for speculative-graded bond is much higher than that for investment-graded bond, which is consistent with the belief that speculative-grade debt is more sensitive to business condition and more likely to default.

A S&P issuer’s rating can be changed to other ratings at any time due to a change in its credit quality. Its rating can be upgraded to a higher rating for an improvement in its credit quality; can be downgraded to a lower for a credit deterioration; can become D for any occurrence of its default event; can become N.R. for the lack of its credit quality information.

By S&P’s definition, “An issuer designated N.R. is not rated”. So rating N.R. does not implies the information of credit quality. It is not a rating but a state of not
being rated. The reasons for an issuer with a N.R. rating could be the extinguishing of its entire debt or S&P’s lack of sufficient information to evaluate its credit quality. Once these reasons are absent, a N.R.-rated issuer could be rated back to any one of the seven initial ratings. N.R. rating is not a terminal state. The fact that S&P does not keep track of the rates of transitions out of N.R. raises some difficulties in modeling rating transition probabilities. We emphasize this issue again in Chapter 6.

Another kind of interesting transition is D (default) transition. Whenever an issuer defaults, its rating is downgraded to D immediately. Subsequently, this issuer’s rating will always be staying at D. So D is different from other ratings in the sense of its rating being an absorbing or terminal property.

S&P’s observed one-year rating transition matrix is composed of the observed transition rates from each of seven initial ratings \{AAA, AA, A, BBB, BB, B, CCC\} to nine final ratings \{AAA, AA, A, BBB, BB, B, CCC, D, N.R.\}. The transition rate of initial rating \(i\) to final rating \(j\) for year \(t\) is calculated by dividing the number of issuers whose rating is transited from \(i\) at the \(t\)th year-beginning to \(j\) at the \(t\)th year-end by the number of initially \(i\)-rated issuers at the year-beginning. A \(7 \times 9\) one-year transition matrix has seven unitary row summations. The transition matrix for any period of time can be constructed in the same way with a different time horizon.

From the above definition of transition rate, a transition matrix involves only the rating observations at two discrete time points - the beginning and the ending of a specific time period like a year. Any other intermediate rating within that period of time will be immaterial. For example, suppose that an issuer had rating \(i\) at the beginning of a year. In the middle of this year its credit rating is changed to rating
and then before the year-end the intermediate rating $j'$ is changed to rating $j$. Because of the discreteness of observation, only the rating change $i$ to $j$ is recorded and intermediate rating $j'$ is ignored. So this inherent observation discreteness could leave out some continuous rating change information. Both the time of transition occurrences and the time horizon for measuring credit risk are continuous variables. Therefore, a continuous model of rating transitions would be much more practical and realistic. In Chapter 5, we further discuss how to adjust the model generated by discrete observations into continuous time framework.

### 4.1.2 Macro-economic time series

#### 1. Raw data description

An issuer’s credit quality is dependent on market conditions according to Nickell et al (2000) [36] and Carty (1997) [17]. We chose 12 time series to represent economic conditions during the time period 1981-1999. They were collected from the Federal Reserve Economic Data website. These variables were:

- **GDP** Real gross domestic product
  - seasonally adjusted, quarterly reported
- **PAYRL** Total nonfarm payrolls
  - seasonally adjusted, monthly reported
- **UNEMP** Civilian unemployment rate
  - seasonally adjusted, monthly reported, percentage
- **CPI** Consumer price index for all urban consumers

---

all item less food and energy
seasonally adjusted, monthly reported

PPI  Producer price index
finished goods less food and energy
seasonally adjusted, monthly reported

R3M  3-Month treasury bill rate, secondary market
monthly average of business days, discount basis, percentage

R6M  6-Month treasury bill rate, secondary market
monthly average of business days, discount basis, percentage

R1Y  1-Year treasury bill rate, secondary market
monthly average of business days, discount basis, percentage

GS05  5-year treasury bill rate, secondary market
monthly average of business days, percentage

GS10  10-year treasury bill rate, secondary market
monthly average of business days, percentage

GS30  30-year treasury bill rate, secondary market
monthly average of business days, percentage

AAA  Moody’s seasoned Aaa corporate bond yield
monthly average of business days, percentage

BAA  Moody’s seasoned Aaa corporate bond yield
monthly average of business days, percentage

2. Data preparation
The data are collected in different measurement scales and in different time units. Therefore, we need to modify these raw macro-economic data before we employ them in our models.

The first issue was *measurement scale*. We notice that GDP, PAYRL, CPI and PPI were quantitative variables and the others were all measured in percentages. The magnitude of these quantitative variables gave little information about the growth rate of the economy. We replace these four quantitative variables by the log of the ratio of the current reading to the previous observation. For example, let $x_t$ denote the raw observation of GDP at time $t$. It was transformed to $\log \frac{x_t}{x_{t-1}}$ in modified time series. For GDP and PAYRL, a positive modification value indicates that the economy was in recovery or expansion while a negative modification value might indicate that the economy was in slow-down or recession. Large positive CPI and PPI are more likely to indicate inflation and large negative CPI and PPI are more likely to indicate deflation.

Next, we *centered* all data at their sample means. That is, at time $t$, the average $\bar{x}$ was subtracted from observation $x_t$ over the 1981-1999 period. The sign of these centered observations can tell the relative comparison of the raw observation to the historically average level.

Moody’s Aaa and Baa corporate bond yields measured the same market risk but different amounts of credit risk. Aaa-rated bonds are considered to be close to risk-free. The yield difference between Aaa-rated and Baa-rated bonds could be considered as a measurement of Baa-bond credit risk. This difference variable was called *SPREAD* in this study. However, SPREAD is only a rough credit risk measurement.
for two reasons. Firstly, there are various times to maturity for bonds in the same rating pool. Time to maturity can be a very important factor to affect bond yield or price in a financial market. To some degree, the lack of this information could generate a different picture of credit risk. Secondly, this SPREAD might over-value credit risk for higher-than-BBB ratings and under-value it for lower-than-BBB ratings.

The next adjustment was to unify the time intervals for all time series. Some raw macro-economic time series were reported quarterly while some were observed monthly. Later, for discrete modeling we require all observations be measured in years and for continuous modeling we require that all observations be recorded in months. Yearly data needed by discrete models are simply the average of monthly or quarterly data within the corresponding year. Most economic variables were observed monthly, which already fits the need of continuous models. The problem was those quarterly variables GDP, PAYRL, CPI and PPI. To convert them into monthly data, we valued each monthly observation within a quarter as that quarterly observation.

To make the data consistent in magnitude, we brought all the modified data into same interval \((-10, 10)\). Specifically, this was achieved by multiplying the adjusted GDP time series by 100 and PAYRL, CPI and PPI by 1000.

Finally, we obtained two sets of 12 modified macro-economic time series for 1981-1999 - one for discrete modeling and the other for continuous modeling. These time series were GDP, PAYRL, CPI, PPI, UNEMP, R3M, R6M, R1Y, GS05, GS10, GS30 and SPREAD. In the following section, \(x_t\) denotes any modified time series. The time variable, \(t = 1, 2, 3 \ldots 19\) years in discrete models and \(t = 1, 2, 3 \ldots 228\) months in continuous models.
4.2 Fundamental assumptions

In this section, we make four assumptions on credit rating transitions over single-period and multi-period times. These assumptions are crucial foundations for our entire credit risk modeling study. Through these assumptions, we incorporate two major concerns of rating migration analysis, business condition dependence and credit risk correlation between two issuers, into our modeling process. We will see their importance as we model and estimate discrete and continuous rating transition probabilities in the following two chapters.

4.2.1 Assumption 1 - identical transition distribution for issuers with same initial rating

An implication in composing single-period transition matrices to summarize credit rating migrations is that all the same-rated issuers have identical transition probabilities over same time horizon. This is also the underlying rationale on which credit agencies base their adoption of categorical credit risk measurement. We keep this assumption in our study.

Suppose at the beginning of year $t$, there are $N_t$ initially rated issuers. Let $r_{k,t}$ denote the $k$th ($1 \leq k \leq N_t$) issuer’s rating at year-beginning and let $r_{k,t+1}$ denote its rating at year-end. Notice $r_{k,t+1}$ should be equivalent to its initial rating for year $t+1$. Because each row summation is equal to one in a transition matrix, the probability of its rating transition to $j$ from $i$ in year $t$ is actually the *conditional* transition
probability to rating $j$ given the known rating $i$. That is,

$$p_{k,t,i,j} = Pr(r_{k,t+1} = j | r_{k,t} = i).$$

where $p_{k,t,i,j}$ is the transition probability of the $k$th issuer from rating $i$ to $j$ in year $t$. With Assumption 1, for any issuer $k$ satisfying $r_{k,t} = i$,

$$p_{k,t,i,j} = p_{t,i,j},$$

and $\sum_j p_{t,i,j} = 1$. The above expression indicates that an issuer’s single-period transition distribution depends on time $t$ and its initial rating $i$, not on issuer $k$ itself. In other words, transition probabilities are issuer-independent.

### 4.2.2 Assumption 2 - business dependence of transition distribution

In this study, we represent the time dependence of single-period transition probability $p_{t,i,j}$ by its dependence on business conditions, which are varying with time. For period $t$, denote the known business condition vector as $\mathbf{x}_t$. Then,

$$p_{t,i,j} = Pr(r_{t+1} = j | r_t = i) = f_{i,j}(\mathbf{x}_t),$$

(4.1)

where $f_{i,j}(\mathbf{x}_t)$ is a function of business condition $\mathbf{x}_t$ at time $t$. The unitary summation constraints remain true: $\sum_{j=1}^g f_{i,j}(\mathbf{x}_t) = 1$, for any $\mathbf{x}_t$ and initial rating $i$.

Because of the variability in business condition time series, for the same rating transitions, the transition probabilities are more likely to be different for different discrete single-periods. For example, let us consider a certain transition from rating
i to j and two different years \( t_1, t_2, t_1 \neq t_2 \). Then these two transition probabilities \( p_{t_1,i,j} \) and \( p_{t_2,i,j} \) will not be the same if \( \mathbf{x}_{t_1} \neq \mathbf{x}_{t_2} \).

This assumption is one realization of the well known dependence of credit risk on business conditions. The existence of dependence has been detected in our Part I study and supported by Altman and Kao (1992) [8], Carty (1997) [17].

### 4.2.3 Assumption 3 - single-issuer Markov rating transition process

For one issuer, we assume its credit rating transitions over several periods follow a Markov process. That is, given an issuer’s rating history up to time period \( t \), the distribution of its next rating transition in that period is equivalent to the transition distribution conditional on its rating at the beginning of period \( t \).

We denote an issuer’s discrete rating history from time 1 to time \( t \) as \( r_1 = i_1, r_2 = i_2, \ldots, r_{t+1} = i_{t+1} \). We express Assumption 3 as,

\[
Pr(r_{t+1} = i_{t+1}|r_1 = i_1, r_2 = i_2, \ldots, r_t = i_t) = Pr(r_{t+1} = i_{t+1}|r_t = i_t),
\]

(4.2)

and by Assumption 2 expression (4.1),

\[
= f_{i_t,i_{t+1}}(\mathbf{x}_t).
\]

(4.3)

From the discussion of Assumption 3, this Markov transition process may or may not be stationary. The stationarity is contingent upon how the business condition vector \( \mathbf{x}_t \) changes with time.

If transition behavior is independent from business conditions or if business conditions are constant over time, then \( f_{i,j}(\mathbf{x}_t) = f_{i,j} \), for any \( t \), i.e. the transition
probability would be constant over time periods. In that case, the transition process
is a stationary Markov process. In fact, the economy has been observed to follow
some cyclic fluctuations, called business cycles. More commonly, \( f_{i,j}(x_t) \neq f_{i,j}(x_{t_2}) \),
for \( t_1 \neq t_2 \). Then rating transition processes should be time-dependent i.e. non-
stationary Markovian.

From this Markovian transition process assumption (4.3), the probability of given
historical rating path \( r_2 = i_2, r_3 = i_3, \ldots, r_{t+1} = i_{t+1} \) with known start rating \( r_1 = i_1 \)
can be derived as,

\[
Pr(r_2 = i_2, r_3 = i_3, \ldots, r_{t+1} = i_{t+1}|r_1 = i_1) = \prod_{s=1}^{t} f_{i_s, i_{s+1}}(x_s).
\]

Hence, the multi-period transition probability from time 1 to time \( t \) would be,

\[
Pr(r_{t+1} = j|r_1 = i) = \sum_{i_2, i_3, \ldots, i_t} \prod_{s=1}^{t} f_{i_s, i_{s+1}}(x_s).
\]
Let $P_{1,t}$ denote the square transition probability matrix from time 1 to time $t$ with $(i,j)$th element $Pr(r_{t+1} = j | r_1 = i)$. In matrix language, the above equation (4.5) can be rewritten as,

$$P_{1,t} = \prod_{s=1}^{t} P_{s,s+1}. \tag{4.6}$$

$P_{s,s+1}$ is $x_s$ dependent because

$$Pr(r_{s+1} = j | r_s = i) = f_{i,j}(x_s)$$

Therefore, this Markovian transition assumption enables us to obtain the multi-period transition matrix as the product of single-period transition matrices with known business time series. We will emphasize this Markov assumption again in Chapter 6 to model the transition probability over any continuous time horizon.

### 4.2.4 Assumption 4 - multi-issuer independent transition distributions

We assume transition distributions for different issuers are independent. Given this assumption, for any two issuers whose ratings are denoted by $r$ and $r'$, their rating transitions to ratings $j$, $j'$ from initial ratings $i, i'$ in year $t$ and year $t'$, respectively, the joint transition probability would be,

$$Pr(r_{t+1} = j, r'_{t'+1} = j'| r_t = i, r'_{t'} = i')$$

$$= Pr(r_{t+1} = j | r_t = i, r'_{t'+1} = j'| r'_{t'} = i')$$

$$= Pr(r_{t+1} = j | r_t = i) \times Pr(r'_{t'+1} = j'| r'_{t'} = i')$$

$$= p_{t,i,j} \times p_{t',i',j'}$$

$$= f_{i,j}(x_t) \times f_{i',j'}(x_{t'})$$
Thus, a joint transition probability is simply the product of individual conditional transition probabilities.

If \( t = t' \), then the rating transition correlation between two issuer’s credit risks is considered through the joint transition probability as a function of known business conditions. That reflects what has been found by Carty (1997) [17]: credit risk correlation varies with the business cycle. Along with Assumption 2 and Assumption 3, instead of modeling credit risk correlation, we can model two issuers’ joint default probability under some known business conditions.

Under the situation \( t \neq t' \) in the above expression, any two issuers’ Markov transition processes are independent by Assumption 3. Moreover, for the joint probability of any two issuers’ rating paths, say, \( r_2 = i_2, r_3 = i_3, \ldots, r_{t+1} = i_{t+1} \) and \( r'_2 = i'_2, r'_3 = i'_3, \ldots, r'_{t+1} = i'_{t+1} \), with known \( r_1 = i_1 \) and \( r'_1 = i'_1 \), we have

\[
Pr(r_2 = i_2, r_3 = i_3, \ldots, r_{t+1} = i_{t+1} | r_1 = i_1; r'_2 = i'_2, r'_3 = i'_3, \ldots, r'_{t+1} = i'_1) = Pr(r_2 = i_2, r_3 = i_3, \ldots, r_{t+1} = i_{t+1} | r_1 = i_1) \\
\times Pr(r'_2 = i'_2, r'_3 = i'_3, \ldots, r'_{t+1} = i'_{t+1} | r'_1 = i'_1),
\]

by equation (4.4),

\[
= \prod_{s=1}^{t} f_{i_s, i_{s+1}}(x_s) \times \prod_{s=1}^{t} f_{i'_s, i'_{s+1}}(x_s).
\]

(4.7)

Assumption 4 is very useful for us to obtain the likelihood function of all observed transitions during 1981 – 1999. We will specify how to come up with the likelihood function in Section 5.1.
4.2.5 Consideration of business conditions

The above discussions are all based on known business conditions \( x_t \) at time \( t \). That is, we simply assume a fixed business conditions time series vector \( x_t \). However, it is more general to treat business condition as random variables. It is especially practical to incorporate the distribution structure of business conditions in predicting future rating transitions.

If we think of \( x_t \) as a stochastic process rather than a known observable, some of the preceding assumption formulae need to be revised to reflect this randomness of business conditions. In other words, we consider the conditional transition probability on business variables instead of modeling the transition probability as a function of known business conditions. For transition from rating \( i \) to \( j \) in year \( t \), the transition probability conditioned on the business variable \( x_t \) is denoted by \( p_{i,j|t} \), and

\[
p_{i,j|x_t} = \Pr(r_{t+1} = j | r_t = i, x_t) \tag{4.8}
\]

The unconditional transition probabilities thus become weighted averages over stochastic business condition distribution space, and the weights are the distribution probabilities of random business conditions. For a simple example, at any time \( t \), let \( x_t \) have two possible states \( v \) or \( w \). Let \( \Pr(x_t = v) + \Pr(x_t = w) = 1 \). Then the unconditional transition probability in period \( t \) from rating \( i \) to \( j \) would be,

\[
p_{t,i,j} = \Pr(r_{t+1} = j | r_t = i) = \Pr(r_{t+1} = j | r_t = i, x_t = v) \Pr(x_t = v) + \Pr(r_{t+1} = j | r_t = i, x_t = w) \Pr(x_t = w),
\]
by equation (4.8), this becomes

\[ p_{i,j|v}Pr(x_t = v) + p_{i,j|w}Pr(x_t = w). \]

In this study, our goal is to propose a discrete and a continuous business conditioned rating transition model. So we consider \( x_t \) as known observations. The transition predicting task, which involves more randomness in business conditions, is left for our future work.
Chapter 5

Discrete Transition Probability Models

In this Chapter, we propose a GLM (generalized linear model) of discrete transition probabilities and describe how ML (maximum likelihood) estimation method was implemented numerically to obtain the estimates of parameters in the discrete model. We will discuss the existence of over-dispersion in our multinomial modeling (which might be caused by unknown random effects) in the last Section. All the discrete models were built over a one-year discrete time horizon.

Below are some basic notations we will need in the Discrete transition modeling.

\[ t \text{ – year } 1, 2, \ldots, 19 \text{ starting with year 1981 and ending with year 1999} \]

\[ i \text{ – initial rating } 1, 2, \ldots, 7 \text{ at the beginning of each year} \]

\[ ^{1}\text{They are numbered in an increasing order from best rating AAA with } i = 1 \text{ to worst rating CCC with } i = 7. \]
$j$ – final rating 1, 2, ..., 9 at the end of each year \(^2\)

$n_{t,i,j}$ – the observed number of issuers with initial rating $i$ and final rating $j$ in year $t$

$x_t$ – the $t$th year modified observation of a business time series

### 5.1 One-year discrete transition models

Let $p_{t,i,j}$ denote the transition probability from rating $i$ to rating $j$ in year $t$. We model the dependence of the rating transition probability $p_{t,i,j}$ on known business covariates through a GLM (generalized linear model) which has a logarithm link between $p_{t,i,j}$ and a linear predictor $\eta_{t,i,j}$. That is,

$$\eta_{t,i,j} = \log p_{t,i,j}$$

for $t = 1, 2, \ldots, 19; i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9$. \hspace{1cm} (5.1)

Because of the restrictions $\sum_j p_{t,i,j} = 1$ for any $t, i$, the inverse transformation can be written in the form

$$p_{t,i,j} = \frac{\exp \eta_{t,i,j}}{\sum_{j'=1}^{9} \exp \eta_{t,i,j'}}.$$ \hspace{1cm} (5.2)

Under the Assumption 2, we modeled $\eta_{t,i,j}$ as a linear function of $L$-dimension business covariate vector $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{L,t})$ at time $t$.

$$\eta_{t,i,j} = \log \pi_{i,j} + \sum_{l=1}^{L} \beta_l x_{l,t},$$ \hspace{1cm} (5.3)

where $\log \pi_{i,j}$ with $\pi_{i,j} > 0$ is constant for rating transitions from $i$ to $j$; $x_{l,t}$ is the observation of the $l$th business covariate in year $t$; $\beta_l$ is the $l$th covariate coefficient which is proportional to the change in the logarithm of transition probability by a

\(^2\)Same rating definitions as $i$ for $j = 1, 2, \ldots, 7; j = 8$ for D rating; and $j = 9$ for N.R. rating.
one unit change in $x_{i,t}$. Obviously, the log transition probability is first modeled as the summation of the transition-related factor $\pi_{i,j}$ and business-related factors $\beta_l x_{l,t}$.

If we do not consider the dependence of rating transitions on business condition, then $\beta_l = 0$ for $k = 1, 2, \ldots, K$, and the above model simply becomes

$$\eta_{t,i,j} = \log\pi_{i,j},$$

(5.4)

with $\beta_l = 0$ for any $k = 1, 2, \ldots, K$. By the logarithm link (5.1), we have

$$p_{t,i,j} = \pi_{i,j}.$$  

(5.5)

A problem in the above GLM modeling (5.3) is the nonestimability of the business condition effect. This is caused by the equivalence between model (5.3) and model (5.4) which is only transition-related. If we use the inverse transformation expression (5.2), these two models both produce expression (5.5). Therefore, we need to further modify the above model so that it can incorporate the impact of known business conditions on transitions.

The GLM in (5.3) is adjusted to be

$$\eta_{t,i,j} = \log\pi_{i,j} + \sum_{l=1}^{L} \beta_l x_{l,t} m_{l,i,j},$$

(5.6)

where $m_{l,i,j}$ represents a business condition $x_t$ effect on transition probability from rating $i$ to $j$. Accordingly, the inverse transformation (5.2) is

$$p_{t,i,j} = \frac{\pi_{i,j} \exp\left(\sum_{l=1}^{L} \beta_l x_{l,t} m_{l,i,j}\right)}{\sum_{j'=1}^{9} \pi_{i,j'} \exp\left(\sum_{l=1}^{L} \beta_l x_{l,t} m_{l,i,j'}\right)}, \text{ for any } t, i, j.$$  

(5.7)

For any $i, j$ in above model (5.7), $\pi_{i,j}$ is the baseline transition probability only related to transition properties. The business condition adjustment on baselines is
represented by component \( \exp(\sum_{l=1}^{L} \beta_l x_{l,t} m_{l,i,j}) \). The effect is defined in the \( m_{l,i,j} \) element. We mainly consider two business condition effects on transitions: one is called the *stability-migration* effect and the other is called the *upgrade-downgrade* effect.

Rating stability is the probability of staying at the same initial rating at year’s end. Stability-migration effect is defined as weaker rating stability under an adverse economic condition and stronger rating stability under a friendly economic condition. For example, the last U.S. economic recession happened in the early 90’s. The overall proportion of unchanged ratings dropped far below average during that time.

An issuer’s rating is upgraded/downgraded if its rating is changed to a higher/lower rating according to an improvement/deterioration of its credit quality. The business upgrade-downgrade effect is defined as more upgrades less downgrades in strong economy and less upgrades more downgrades in weak economy. S&P observes that the ratio of downgrades over upgrades, excluding the default rating, stayed very high in the 1990 and 1991 recession years.

We represent the rating stability-migration effect or upgrade-downgrade effect from known business condition by \( m_{l,i,j} \)’s.

For the \( l \) th business covariate \( x_{l,t} \)’s stability-migration effect, we define

\[
m_{l,i,j} = \begin{cases} 
1 & \text{if } j = i, \\
-1 & \text{if } j \neq i,
\end{cases}
\]  

(5.8)

where \( i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9 \). Let \( M_1 \) be the \( 7 \times 9 \) matrix form of the stability-
migration effect representation. Then we have

\[
M_1 = \begin{bmatrix}
1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-1 & -1 & 1 & -1 & -1 & -1 & -1 \\
-1 & -1 & 1 & -1 & -1 & -1 & -1 \\
-1 & -1 & -1 & 1 & -1 & -1 & -1 \\
-1 & 1 & -1 & -1 & 1 & -1 & -1 \\
-1 & -1 & -1 & -1 & 1 & -1 & -1 \\
-1 & -1 & -1 & -1 & -1 & 1 & -1 \\
1 & -1 & -1 & -1 & -1 & -1 & 1
\end{bmatrix}.
\]

For the \(k\)th business covariate \(x_{t,i}'s\) upgrade-downgrade effect, we let

\[
m_{t,i,j} = \begin{cases}
1 & \text{if } j < i, \\
-1 & \text{if } j > i, \\
0 & \text{if } j = i,
\end{cases}
\]

where \(i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 8\). One problem arose here: should we treat the transition to N.R. (\(j = 9\)) as upgrade or as downgrade? As we mentioned in Chapter 3, there are two major reasons for assigning a N.R. rating: debt-extinguish or lack of rating information [13]. Carty [17] finds that the correlation between rating withdrawals and rating changes is very thin. We treat the transitions from investment grades to N.R. as upgrades, i.e. \(m_{i,i,9} = 1\) for \(i = 1, 2, 3, 4\). By contrast, N.R. transitions from speculative grades should be thought of as downgrades, i.e. \(m_{i,i,9} = -1\) for \(i = 5, 6, 7\). In the matrix form, we have
Notice that for any constants \( c \), \( \{\eta_{t,i,j}\} \) and \( \{\eta_{t,i,j} + c\} \) would come up with same set of \( p_{t,i,j} \) by model \((5.7)\). Therefore, we need a constraint on \( \pi_{i,j} \) to make the estimate identifiable: for \( i = 1, 2, \ldots, 7 \),
\[
\sum_{j=1}^{9} \pi_{i,j} = 1.
\]

**5.1.1 Null transition model**

We need to test the significance of business condition impact on rating transition. Thus, we set a business-independent model, i.e. null transition model, as a comparison null model for business-dependent models. All “no business condition covariate” model can be derived under the null hypothesis

\[
H_0 : \quad \beta_l = 0 \quad \text{for } l = 1, 2, \ldots, L.
\]

The GLM model \((5.7)\) reduces to

\[
p_{t,i,j} = \frac{\pi_{i,j}}{\sum_{j' = 1}^{9} \pi_{i,j'}}.
\]
because of the constraint $\sum_{j'} \pi_{i,j'} = 1$, $p_{t,i,j}$ is just

$$= \pi_{i,j}.$$ (5.13)

So the probabilities of a given transition are the same in different years. Under this situation, the transition process follows a stationary Markov process using the time-homogeneous transition probabilities.

### 5.1.2 One-covariate transition model

The single business covariate transition model is the general model (5.7) with one business covariate, that is, the total number of covariates in model (5.3) is $L = 1$. For the one-covariate model, $\beta_i$, $x_{1,t}$ and $m_{1,i,j}$ are replaced by $\beta$, $x_t$ and $m_{i,j}$, separately. The model expression is, for $t = 1, 2, \ldots, 19; i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9$,

$$p_{t,i,j} = \frac{\pi_{i,j} \exp (\beta x_t m_{i,j})}{\sum_{j'}^{9} \pi_{i,j'} \exp (\beta x_t m_{i,j'})},$$ (5.14)

with constraint,

$$\sum_{j'}^{9} \pi_{i,j'} = 1, \text{ for any } i.$$ (5.15)

Here $m_{i,j}$ can be the $(i, j)$th element either in the stability-migration effect representation $M_1$ or the upgrade-downgrade effect representation $M_2$.

### 5.1.3 Two-covariate transition model

Similarly, if we jointly model the effects of two business covariates, then $L = 2$ in the GLM model (5.7). And the model is expressed as, for $t = 1, 2, \ldots, 19; i =$
1, 2, \ldots, 7; j = 1, 2, \ldots, 9,

\[ p_{t,i,j} = \frac{\pi_{i,j} \exp (\beta_1 x_{1,t} m_{1,i,j} + \beta_2 x_{2,t} m_{2,i,j})}{\sum_{j'=1}^{9} \pi_{i,j'} \exp (\beta_1 x_{1,t} m_{1,i,j'} + \beta_2 x_{2,t} m_{2,i,j'})} \]  

(5.16)

with constraint,
\[ \sum_{j'=1}^{9} \pi_{i,j'} = 1. \]

Once again, \( m_{1,i,j} \) and \( m_{2,i,j} \) are the \((i, j)\)th element in stability-migration effect representation \( M_1 \) and upgrade-downgrade effect representation \( M_2 \), respectively, with known business covariates \( x_{1,t}, x_{2,t} \).

Both the single business covariate model and double business covariate model are dependent on the value of known business covariate time series. Because of the variability in macro-economic variables with time, the transitions modeled by these two models are nonstationary Markov processes.

In the next section, we introduce the construction of a log likelihood function and numerically estimate the parameters by ML estimation method.

### 5.2 Maximum likelihood estimation

#### 5.2.1 Likelihood function

In Assumption 3 – Markovian transition process, we derived the probability of a single issuer’s multi-year rating transition trajectory as the product of all its one-year transition probability within that period, see equation (4.4) for details. Moreover, under Assumption 4, we know mutli-issuer’s rating transition histories are independent. Equation (4.7) gives the joint rating transition trajectory for two issuers over
a multi-period time. More generally, the assumed independence among all issuers’
Markovian transition processes enables us to derive the joint transition likelihood for
all observed transitions as the product of every issuer’s rating transition likelihood.

Suppose there are \( N \) rated issuers at the beginning of year 1 and assume every
year we can trace each issuer’s rating transitions from year 1 to year \( T \). Let \( i_{k,t} \) for
\( k = 1, 2, \ldots, N; t = 1, 2, \ldots, T + 1 \) denote the initial rating of the \( k \)th issuer in year \( t \);
notice \( i_{k,T+1} \) is its final rating at the end of year \( T \). Then, the generalization of two
issuers’ joint transition trajectory probability (4.7) would be

\[
Pr(r_{k,t} = i_{k,t}, t = 2, 3, \ldots, T + 1 | r_{k,1} = i_{k,1}; k = 1, 2, \ldots, N)
\]  

(5.17)

by independence Assumption 4, this is

\[
= \prod_{k=1}^{N} Pr(r_{k,t} = i_{k,t}, t = 2, 3, \ldots, T + 1 | r_{k,1} = i_{k,1}),
\]  

(5.18)

and by Markov process Assumption 3, it becomes

\[
= \prod_{k=1}^{N} \prod_{t=1}^{T} Pr(r_{k,t+1} = i_{k,t+1} | r_{k,t} = i_{k,t}),
\]  

(5.19)

\[
= \prod_{k=1}^{N} \prod_{t=1}^{T} p_{t,i_{k,t},i_{k,t+1}},
\]  

(5.20)

then, letting \( n_{t,i,j} \) be the number of issuers in year \( t \) with rating transition to \( j \) from
\( i \), it becomes

\[
= \prod_{t=1}^{T} \prod_{i=1}^{7} \prod_{j=1}^{9} p_{t,i,j}^{n_{t,i,j}},
\]  

(5.21)

where \( \sum_{i=1}^{7} \sum_{j=1}^{9} n_{t,i,j} = N \) for any \( t = 1, 2, \ldots, T \).

The above joint transition probability expression (5.21) implies that any rating
transitions are able to be observed. Any transitions among ratings AAA, AA, A,
BBB, BB, B, CCC are recorded in S&P’s one-year transition matrix. But there are two exceptions, namely for transitions to ratings D and N.R..

For any issuer downgraded to D, its rating will stay at D after the downgrade and can not be transited to any other rating after that. The absorbing property of D implies: \( Pr(r_{t+1} = D | r_t = D) = p_{t,9,9} = 1 \) for any \( t \), so that D transitions bring zero uncertainty into the likelihood.

It is true that issuers with N.R. can be rated back to any other rating. However, lack of N.R. observations in S&P’s one-year transition report led us to assume that N.R. is also an absorbing state. That is, \( Pr(r_{t+1} = N.R. | r_t = N.R.) = p_{t,8,8} = 1 \). In the next Chapter on continuous modeling, we will keep this assumption to produce a square modification of the transition matrices.

We express the likelihood for all transition observations in the time period between year 1981 and year 1999 as

\[
L = \prod_{t=1}^{19} \prod_{i=1}^{7} \prod_{j=1}^{9} p_{t,i,j}^{n_{t,i,j}}. \tag{5.22}
\]

where \( n_{t,i,j} \geq 0 \) and \( \sum_{j=1}^{9} p_{t,i,j} = 1 \) for any \( t, i \). Because raising any number to the zero power gives one, it is equivalent to restrict \( n_{t,i,j} > 0 \). This restriction will be useful to derive a log likelihood function for conditional models.

The log likelihood function is

\[
\log L = \sum_{t=1}^{19} \sum_{i=1}^{7} \sum_{j=1}^{9} n_{t,i,j} \log p_{t,i,j}, \quad \text{for} \quad n_{t,i,j} > 0 \tag{5.23}
\]

by GLM (5.6), it is

\[
= \sum_{t=1}^{19} \sum_{i=1}^{7} \sum_{j=1}^{9} n_{t,i,j} \left( \log \pi_{i,j} + \sum_{l=1}^{L} \beta_{l} x_{t,i,m_{l,i,j}} \right). \tag{5.24}
\]
This log likelihood function is not defined in the restrict sense because every year S&P has more new issuers added into the rated issuer pool, that is, \( \sum_{i=1}^{7} \sum_{j=1}^{9} n_{t,i,j} \) is not constant for different years. In the above log likelihood function (5.24), the varying numbers of issuers over time contribute different weights for different year transition probabilities.

Specifically, the three discrete models have log likelihood functions as below. For the null transition model,

\[
\log L_0 = \sum_t \sum_i \sum_j n_{t,i,j} \log \pi_{i,j}; \tag{5.25}
\]

for the one-covariate model,

\[
\log L_1 = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ \log \pi_{i,j} + \beta x_{t,m_{i,j}} - \log \left[ \sum_{j' = 1}^{9} \pi_{i,j'} \exp (\beta x_{t,m_{i,j'}}) \right] \right\}; \tag{5.26}
\]

for the two-covariate model,

\[
\log L_2 = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ \log \pi_{i,j} + \beta_1 x_{1,t} m_{1,i,j} + \beta_2 x_{2,t} m_{2,i,j} - \log \left[ \sum_{j' = 1}^{9} \pi_{i,j'} \exp (\beta_1 x_{1,t} m_{1,i,j'} + \beta_2 x_{2,t} m_{2,i,j'}) \right] \right\}. \tag{5.27}
\]

The method of maximum likelihood estimation is adopted to obtain the estimates of parameters: base-line transition probabilities \( \pi_{i,j} \) and covariate coefficients \( \beta, \beta_1, \beta_2 \). The Maximum Likelihood Estimate (MLE) is the parameter point for which the observed sample is most likely in that it maximizes the likelihood function. The uniqueness of ML estimation for multinomial log linear models has been proved by Agresti [4] see page 455. Our task was to find these ML estimates by numerically
solving the log likelihood equations, i.e.

\[ \frac{\partial \log L(\theta)}{\partial \theta} = 0, \quad (5.28) \]

where \( \theta \) is the parameter vector. Specifically,

for the null model,

\[ \theta (\pi_{i,j}, i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9; i \neq j); \quad (5.29) \]

for the one-covariate model,

\[ \theta (\beta, \pi_{i,j}, i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9; i \neq j); \quad (5.30) \]

for the two-covariate model,

\[ \theta (\beta_1, \beta_2, \pi_{i,j}, i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9; i \neq j). \quad (5.31) \]

For the null model without any business covariate, the zero first differentiation equation (5.28) becomes

\[ \frac{\partial \log L_0}{\partial \pi_{i,j}} = 0 \quad \text{for } i = 1, 2, \ldots, 7; \ j = 1, 2, \ldots, 9; \quad (5.32) \]

with constraint

\[ \sum_{j=1}^{9} \pi_{i,j} = 1 \quad \text{for } i = 1, 2, \ldots, 9. \quad (5.33) \]

The solution to this first derivative equation is

\[ \hat{\pi}_{i,j} = \frac{n_{i,j}}{n_{i,.}}, \quad (5.34) \]

where \( n_{i,j} = \sum_t n_{t,i,j} \) and \( n_{i,.} = \sum_t \sum_j n_{t,i,j} \).
For a one-covariate model, the first differential functions w.r.t. the covariate coefficient $\beta$ is

$$\frac{\partial \log L_1}{\partial \beta} = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ x_{t,ij} - \frac{\sum_{j'} \hat{\pi}_{i,j'} \exp (\beta x_{t,j'} m_{i,j'}) x_{t,i,j'}}{\sum_{j'} \hat{\pi}_{i,j'} \exp (\beta x_{t,j'} m_{i,j'})} \right\}; \quad (5.35)$$

and the first derivative of the log likelihood function w.r.t. baseline $\pi_{i,j}$ is

$$\frac{\partial \log L_1}{\partial \pi_{i,j}} = \frac{\sum_t n_{t,i,j}}{\pi_{i,j}} - \sum_t \sum_j n_{t,i,j} \frac{\exp (\beta x_{t,m_{i,j}})}{\sum_{j'} \pi_{i,j'} \exp (\beta x_{t,m_{i,j'}})}$$

for any $i, j = \frac{n_{i,j}}{\pi_{i,j}} - \sum_t \sum_{j'} \frac{n_{t,i,j'} \exp (\beta x_{t,m_{i,j'}})}{\sum_{j'} \pi_{i,j'} \exp (\beta x_{t,m_{i,j'}})}$

under same constraint (5.33).

We equate the above first derivatives to zero, and find the MLEs satisfy equations (5.36) and (5.37).

$$0 = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ x_{t,i,j} - \frac{\sum_{j'} \hat{\pi}_{i,j'} x_{t,j'} m_{i,j'}}{\sum_{j'} \hat{\pi}_{i,j'} \exp (\beta x_{t,j'} m_{i,j'})} \right\}; \quad (5.36)$$

$$\hat{\pi}_{i,j} = \frac{n_{i,j}}{\sum_t \sum_{j'} \frac{n_{t,i,j'} \exp (\beta x_{t,m_{i,j'}})}{\sum_{j'} \pi_{i,j'} \exp (\beta x_{t,m_{i,j'}})}}$$

for all $i \neq j. \quad (5.37)$

Because of above functions’ nonlinearity and complexity, we can not find the MLEs of unknown parameters analytically. We need to use some numerical optimization method to obtain the solutions.

For the two-covariate model, the first derivatives w.r.t $\beta_1$, $\beta_2$ and $\pi_{i,j}$ are

$$\frac{\partial \log L_2}{\partial \beta_1} = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ x_{1,t,m_{1,i,j}} \right. \left. - \frac{\sum_{j'} \pi_{i,j'} \exp (\beta_1 x_{1,t,m_{1,i,j'}} + \beta_2 x_{2,t,m_{2,i,j'}}) x_{1,t,m_{1,i,j'}}}{\sum_{j'} \pi_{i,j'} \exp (\beta_1 x_{1,t,m_{1,i,j'}} + \beta_2 x_{2,t,m_{2,i,j'}})} \right\};$$

$$\frac{\partial \log L_2}{\partial \beta_2} = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ x_{2,t,m_{2,i,j}} \right. \left. - \frac{\sum_{j'} \pi_{i,j'} \exp (\beta_1 x_{1,t,m_{1,i,j'}} + \beta_2 x_{2,t,m_{2,i,j'}}) x_{2,t,m_{2,i,j'}}}{\sum_{j'} \pi_{i,j'} \exp (\beta_1 x_{1,t,m_{1,i,j'}} + \beta_2 x_{2,t,m_{2,i,j'}})} \right\};$$
\[ \frac{\partial \log L_2}{\partial \pi_{i,j}} = \frac{\sum_t n_{t,i,j}}{\pi_{i,j}} - \sum_t n_{t,i,j} \exp \left( \beta_1 x_{1,t} m_{1,i,j} + \beta_2 x_{2,t} m_{2,i,j} \right) \sum_{t',j} \sum_{i,j'} \exp \left( \beta_1 x_{1,t'} m_{1,i,j'} + \beta_2 x_{2,t'} m_{2,i,j'} \right) \]

\[ = \frac{n_{t,i,j}}{\pi_{i,j}} - \frac{\sum_t n_{t,i,j} \exp \left( \beta_1 x_{1,t} m_{1,i,j} + \beta_2 x_{2,t} m_{2,i,j} \right)}{\sum_{t'} \pi_{i,j'} \exp \left( \beta_1 x_{1,t'} m_{1,i,j'} + \beta_2 x_{2,t'} m_{2,i,j'} \right)} \quad \text{for any } i, j. \]

Similarly, the MLE could be derived from

\[ 0 = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ x_{1,t} m_{1,i,j} - \frac{\sum_{t'} \hat{\pi}_{i,j'} \exp \left( \hat{\beta}_1 x_{1,t} m_{1,i,j'} + \hat{\beta}_2 x_{2,t} m_{2,i,j'} \right) x_{1,t} m_{1,i,j'}}{\sum_{t'} \hat{\pi}_{i,j'} \exp \left( \hat{\beta}_1 x_{1,t} m_{1,i,j'} + \hat{\beta}_2 x_{2,t} m_{2,i,j'} \right)} \right\}; \quad (5.38) \]

\[ 0 = \sum_t \sum_i \sum_j n_{t,i,j} \left\{ x_{2,t} m_{2,i,j} - \frac{\sum_{t'} \hat{\pi}_{i,j'} \exp \left( \hat{\beta}_1 x_{1,t} m_{1,i,j'} + \hat{\beta}_2 x_{2,t} m_{2,i,j'} \right) x_{2,t} m_{2,i,j'}}{\sum_{t'} \hat{\pi}_{i,j'} \exp \left( \hat{\beta}_1 x_{1,t} m_{1,i,j'} + \hat{\beta}_2 x_{2,t} m_{2,i,j'} \right)} \right\}; \quad (5.39) \]

subsequently,

\[ \hat{\pi}_{i,j} = \frac{n_{t,i,j}}{\sum_t \sum_{t'} \hat{\pi}_{i,j'} \exp \left( \hat{\beta}_1 x_{1,t} m_{1,i,j'} + \hat{\beta}_2 x_{2,t} m_{2,i,j'} \right)}, \quad \text{for } j \neq i. \quad (5.40) \]

Even though [4] shows the uniqueness and convergence of the MLE for GLM models, the complexity of the first derivatives in the one-covariate model and the two-covariate model makes the analytically estimating the MLE impossible. Thus, we use a numerical optimization method to obtain the solutions.

### 5.2.2 Maximum likelihood estimation implementation

The number of unknowns in our models is quite large, which generates some difficulties in computation and estimation. CDO (cyclic descent optimization) is an optimization algorithm that seeks to optimize a multivariate function by optimizing each coordinate (keeping the other coordinates fixed) with a univariate optimization...
technique. This univariate optimization is repeated until a fixed point is reached.

GCDO is the generalized CDO in the sense that it breaks down a parameter vector into smaller sub-vectors which may or may not be univariates.

1. One-covariate model

From equation (5.36), we noticed that the MLE \( \hat{\beta} \) is a function of all \( \hat{\pi}_{i,j} \); and equation (5.37) shows that \( \hat{\pi}_{i,j} \) is dependent on \( \hat{\pi}_{i,j'} \) for all \( j' \) and \( \hat{\beta} \). The parameters are broken into two sets: one set only includes \( \beta \) and the other includes all \( \pi_{i,j} \). The GCDO optimization can be carried out between these two subsets by alternatively estimating one set of parameters and fixing the other with the most updated optimizer. The iteration procedure stops when all the estimates converge.

We started with some fixed set of \( \pi_{i,j} \), \( i = 1, 2, \ldots, 7; \ j = 1, 2, \ldots, 9 \). The log likelihood function (5.26) thus has only one more unknown parameter \( \beta \). From the first derivative function (5.35), we derive the second derivative function as

\[
\frac{\partial^2 \log L_1}{\partial \beta^2} = -\sum_t \sum_i \sum_j n_{t,i,j} \left[ \frac{n_{t,i,j}}{\sum_{j'} \pi_{i,j'} \exp (\beta x_t m_{i,j'})} \right]^2 \times \left\{ \left[ \sum_{j'} \pi_{i,j'} x_t^2 m_{i,j'}^2 \exp (\beta x_t m_{i,j'}) \right] \left[ \sum_{j''} \pi_{i,j''} \exp (\beta x_t m_{i,j''}) \right] - \left[ \sum_{j'} \pi_{i,j'} x_t m_{i,j'} \exp (\beta x_t m_{i,j'}) \right]^2 \right\} \tag{5.41}
\]

Since the first two derivative functions are continuous, the function is approximately parabolic near the maximum \( \beta \) if it exists. We adopted Brent’s method to
maximize this smooth log likelihood function w.r.t $\beta$ for any fixed set of $\hat{\pi}_{i,j}$, $i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9$.

Once the optimum $\hat{\beta}$ is obtained, $\pi_{i,j}$ for $i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9$ need to be estimated by maximizing the likelihood function with this updated estimated $\hat{\beta}$. If there exists such a set of $\hat{\pi}_{i,j}$ that maximizes the likelihood function. Then the first derivatives w.r.t $\pi_{i,j}$ would be all zero. For small $\beta$, $\exp (\beta x_t m_{i,j})$ is close to one for any $i, j$. And the denominator $\sum_{j'} \pi_{i,j'} \exp (\beta x_t m_{i,j'})$ would not be boundless because of the constraint $\sum_{j'} \pi_{i,j'} = 1$. Hence, the adjustment factor $\frac{\exp (\beta x_t m_{i,j'})}{\sum_{j'} \pi_{i,j'} \exp (\beta x_t m_{i,j'})}$ would not change the transition probability $p_{t,i,j}$ from baseline $\pi_{i,j}$ too much.

There are several considerations regarding the numerical approximation. Firstly, we need to set a numerical convergence criterion — a tolerance. If the difference between the current and the last values of the log likelihood function is smaller than this tolerance, then the log likelihood function is assumed to have reached its convergence at the current estimate. In this study, tolerance $tol = 10^{-9}$.

Our second concern is the choice of starting values for baselines $\pi_{i,j}$ and covariate coefficient $\beta$. We believe that rating migrations are primarily affected by the inherent rating transition factor $\pi_{i,j}$ and only slightly adjusted by business conditions $\beta x_t m_{i,j}$. So the MLE $\hat{\beta}$ is expected to be small, which is actually verified by some preliminary optimization study. Therefore, we choose as a starting value, $\beta = 0$ and hence the starting values for $\pi_{i,j}$ are the same as in the null model MLE (5.34).

Another criterion is the upper limit for number of iterations for both univariate Brent’s optimization algorithm with fixed $\beta$ and the multivariate $\pi_{i,j}$ approximation with fixed $\beta$. This maximal iteration number is $max = 1000$. Hence, the optimization
procedure could stop either because of reaching the convergence criteria \( tol \) or because of the large number of iterations. In the result chapter, we found that all the estimates converge for the log likelihood function.

Finally, the bracket for Brent’s estimation of \( \beta \) is set to be the interval \((-0.7, 0.7)\). Below is the detailed algorithm to get the ML estimates of \( \beta \) and \( \pi_{i,j} \).

1. Set the starting value for the business covariate coefficient \( \hat{\beta} = 0 \) and baselines \( \hat{\pi}_{i,j} = \frac{n_{i,j}}{n_{i,i}} \). Set the CDO iteration counter to zero: \( \text{counter} = 0; \)

2. Use Brent’s numerical optimization method to maximize log likelihood function \( \log L_1 \) for fixed baselines; Denote the maximum of the log likelihood function as \( l_1 \) and denote the MLE of \( \beta \) as \( \hat{\beta}; \)

3. Iteratively approximate baseline values \( \pi_{i,j} \) for the last estimated \( \hat{\beta} \) using derivative equation (5.37) until the estimates of \( \pi_{i,j} \) converge or the iteration number reaches the upper boundary \( max \). Denote the corresponding log likelihood function value as \( l_2 \) and final estimates of \( \pi_{i,j} \) as \( \hat{\pi}_{i,j}; \)

4. \( \text{counter}++; \)

5. Define \( \text{diff} = l_2 - l_1 \). Obviously, \( \text{diff} \geq 0. \)

- If \( \text{diff} > tol \) and \( \text{counter} < max \), then repeat step 2, 3, 4;

- If \( \text{diff} \leq tol \) or \( \text{counter} = max \), then stop.

We noted that the max in step 3 never is reached and in fact the numerical convergence is obtained after only a few runs.

2. Two-covariate model
As in the one-covariate optimization GCDO, we have similar differential equations to (5.38), (5.39) and (5.40).

Again, the 58 dimension multivariate parameter vector is broken down into two subsets - one has $\beta_1$ and $\beta_2$; the other has baselines $\pi_{i,j}$. We apply GCDO to maximize the log likelihood function $\log L_2$.

A major difference between the two-covariate matrix optimization versus the one-covariate optimization is that the subset of business covariate coefficients is not univariate but bivariate ($\beta_1, \beta_2$). We need another kind of optimization algorithm to estimate bivariate ($\beta_1, \beta_2$). Powell’s “direction set” method is used for this bivariate optimization problem.

The $tol, max$ and starting values for $\pi_{i,j}$ are set in the same way as those in the one-covariate model.

The detailed numerical algorithm is:

1. Set starting value for business covariate coefficient $\hat{\beta} = 0$ and baselines $\hat{\pi}_{i,j} = \frac{n_{i,.} \cdot n_{j,.}}{n_{i,j}}$. Set the CDO iteration counter to zero: $counter = 0$;

2. Use Powell’s “direction set” method to maximize $\log L_2$ with fixed baselines;
   Denote the maximum of the log likelihood function by $l_1$ and denote the ML estimates by $\hat{\beta}_1, \hat{\beta}_2$;

3. Iteratively approximate the baseline $\pi_{i,j}$ for the most recent estimated $\hat{\beta}_1, \hat{\beta}_2$ using the first derivative equation (5.40) until the estimates of $\pi_{i,j}$ converge or the iteration number reaches the upper boundary $max$. Denote the corresponding log likelihood function value by $l_2$ and final estimates of $\pi_{i,j}$ by $\hat{\pi}_{i,j}$;
4. counter++;

5. Define \( \text{diff} = l_2 - l_1 \). Obviously, \( \text{diff} \geq 0 \).
   
   - If \( \text{diff} > \text{tol} \) and \( \text{counter} < \text{max} \), then repeat step 2, 3, 4;
   
   - If \( \text{diff} \leq \text{tol} \) or \( \text{counter} = \text{max} \), then stop.

5.3 Goodness-of-fit test, over-dispersion and scaled-deviance

5.3.1 Goodness-of-fit \( \chi^2 \) test

Given the 19 observed one-year transition matrices, we fitted the null model, one-covariate and two-covariate discrete models. The question then arose how to measure the discrepancy between the fitted values and corresponding observations. One common approach used for the multinomial distribution is Pearson \( \chi^2 \) statistics defined in [4]. It has the form

\[
X^2 = \sum_l \sum_i \sum_j n_{t,i,j} \left( \frac{\hat{p}_{t,i,j} - \tilde{p}_{t,i,j}}{\hat{p}_{t,i,j}} \right)^2;
\]

where \( \tilde{p}_{t,i,j} = \frac{n_{t,i,j}}{n_{t,i}} \) is the observed transition rate and \( \hat{p}_{t,i,j} \) is the ML estimated transition probability.

Notice there are five transitions that have never been observed historically: AAA→B, AAA→CCC, AAA→D, B→AAA and C→AA. These zero occurrences reduce the number of baseline estimates by five. The number of parameters (no.) is 51, 52 and 53 for the null model, the one-covariate model and the two-coavriate model, re-
spectively. Because we assumed the transition in 19 years for seven initial ratings were conditionally independent, the total number of different transitions would be $19 \times 3 \times 9 = 1197$. Moreover, there are $19 \times 7 = 133$ unitary summation constraints. Therefore, \(d.f. = 1191-133-no\). For the discrete null model, \(d.f. = 1013\); for the one-covariate model, \(d.f. = 1012\); and for the two-covariate model, \(d.f. = 1011\).

5.3.2 Over-dispersion

In the next Result chapter, we will see that almost all $\chi^2$ statistics are as large as 2.7 times of 95% cutoff $\chi^2$ distribution with same \(d.f.\), no matter which model and which covariate time series is applied. These great discrepancies between the fitted values and observations indicates that there is more variation other than simple sampling variation. Firstly, we need to explain the Multinomial sampling distribution assumption for same rating transitions.

The count variable for rating transitions to \(j\) from \(i\) in year \(t\) is \(n_{t,i,j}\), for any \(t, i, j\). Obviously, \(n_{t,i,j}\) is a random nonnegative integer. We assume it has the simplest sampling distribution – a Poisson sampling distribution with mean \(n_{t,i} \cdot p_{t,i,j}\). Then its probability distribution function is

\[
\frac{(n_{t,i} \cdot p_{t,i,j})^{n_{t,i,j}} \exp(-n_{t,i} \cdot p_{t,i,j})}{n_{t,i,j}!} \quad \text{for } n_{t,i,j} = 0, 1, 2, \ldots.
\]

For year \(t\), the total number of initially \(i\)-rated issuers is given as \(n_{t,i,\cdot}\) is given. Therefore, the nine Poisson distributed transition counts \(n_{t,i,j}, j = 1, 2, \ldots, 9\) have Poisson distributions conditional on \(n_{t,i,\cdot} = \sum_{j'=1}^{9} n_{t,i,j'}\). The set of transition counts \(\{n_{t,i,j}\}\), thus, has a multinomial sampling distribution with sample size \(n_{t,i,\cdot}\) and cell probabilities \(\{p_{t,i,j}\}\) under constraint $\sum_{j'=1}^{9} p_{t,i,j} = 1$. 
In this study, there are $19 \times 7 = 153$ multinomial distributions. With the Independence Assumption 4, these 153 multinomial distributions are independent. So, an over-dispersion factor measures the degree to which the variance of the response $n_{t,i,j}$ from our model exceeds the nominal multinomial variance, $n_{t,i}p_{t,i,j}(1 - p_{t,i,j})$.

According to the over-dispersion discussed by McCullagh and Nelder [34], over-dispersion can arise in a number of ways. In this study, we presume these over-dispersions might arise from some unobservable random effects. McCullagh and Nelder (1983) [34] suggests not relying on a specific form of over-dispersion. We define the dispersion factor as $\phi$. That is, the mean of the MLE is assumed unaffected by over-dispersion while the variance is inflated by $\phi$.

More precisely, let $Y$ denote a multinomial distributed vector, $\sim \theta$ denote the vector of probabilities in this multinomial distribution and matrix $\Sigma$ denote the usual multinomial covariance matrix. Suppose the sample size is $m$. According to [34], then we have

$$E(Y) = m\theta;$$
$$cov(Y) = \phi \Sigma.$$ 

The covariance matrix of parameters, obtained from the multinomial log likelihood, is inflated by the dispersion factor $\phi > 1$. Hence, after we fit the data, there remains only the problem of estimating the dispersion factor, which is required for setting confidence limits on parameters. And the over-dispersion $\phi$ can be measured as

$$\phi = \frac{X^2}{\text{residual d.f.}}; \quad (5.43)$$
where $X^2$ is the general Pearson statistic. This estimate is approximately unbiased for $\phi$, is consistent for large sample size, and moreover is approximately independent of the estimated parameters.

### 5.3.3 Scaled-deviance

The discrete null model model, one-covariate model and two-covariate model make up a set of nested models in that each model includes one term more than the previous one. Under the presence of a constant dispersion factor, the first difference of scaled deviance of two nested models is used to test whether the addition of a term was significant or not.

Scaled-deviance refers to the over-dispersion adjusted deviance by [34]. It is obtained by dividing twice the difference between the maximum achievable log likelihood and the fitted likelihood from a model with the estimated over-dispersion factor. The first difference of scaled-deviances from any two nested models is simply twice the difference of the two log likelihood values deflated by the dispersion $\phi$. For instance, model $a$ is nested by model $b$ with one more factor, the reduction of scaled-deviance is

$$D_s = \frac{2(\log L_b - \log L_a)}{\phi}. \quad (5.44)$$

According to McCullagh and Nelder (1989) [34], the $\chi^2$ approximation is usually quite accurate for difference of deviances with one d.f. coming from one additional covariate term in the extended model.

In this study, the interesting scaled-deviance comparisons are those between the null model and the one-covariate model; between the null model and the two-covariate
model; between the two-covariate model and two corresponding one-covariate models.

1. Comparison between null model and one-covariate model

The scaled-deviances difference between the null model and the one-covariate model can help us to test whether one business covariate with certain transition impact could significantly improve the fit from a business-free null model. In the language of statistics, this is equivalent to test

\[ H_{00} : \beta = 0, \]
\[ H_{01} : \beta \neq 0. \]

(5.45)

The reduction in scaled-deviance is calculated as

\[ D_{s1} = \frac{2(\log L_1 - \log L_0)}{\phi}. \]

(5.46)

If \( D_{s1} \) is approximated by a \( \chi^2 \) distribution with 1 d.f., then we can use this statistic to test the significance of a business impact comparing \( D_{s1} \) with some cutoff value from a 1 d.f. \( \chi^2 \) distribution.

Because the reduction in scaled-deviance can be thought of as a quantified measurement of improvement from the null model to the one-covariate model, for 12 business covariate, we have: the greater the \( D_{s1} \), the more important is the impact of the business variables on transition probabilities. That is, scaled-deviance reduction can help us to identify the most significant business variables.

2. Comparison between one-covariate model and two-covariate model

The scaled-deviance difference between a one-covariate model and its nesting two-covariate model allows us to test whether or not adding one more business impact has significant likelihood improvement from the one-covariate model.
First, to test the significance of adding stability-migration impact on one-covariate upgrade-downgrade model, the test hypothesis is

\[ H_{10} : \beta_1 = 0 \text{ and } \beta_2 \neq 0, \]
\[ H_{11} : \beta_1 \neq 0 \text{ and } \beta_2 \neq 0. \]

(5.47)

Second, to test the significance of adding upgrade-downgrade impact on one-covariate stability-migration model, the testing hypothesis is

\[ H_{20} : \beta_1 = 0 \text{ and } \beta_2 = 0, \]
\[ H_{21} : \beta_1 = 0 \text{ and } \beta_2 \neq 0. \]

(5.48)

The statistic of reduction in scaled-deviance is

\[ D_{s_2} = \frac{2(\log L_2 - \log L_1)}{\phi}. \]

(5.49)

This \( D_{s_2} \) has a 1 d.f. \( \chi^2 \) distribution. More specifically, let \( \beta_1 \) be the stability-migration covariate coefficient and \( \beta_2 \) be the upgrade-downgrade covariate coefficient in a two-covariate model. That is, to test hypothesis in (5.47), we need to use

\[ D_{s_{2,1}} = \frac{2(\log L_2 - \log L_{M_2})}{\phi}, \]

(5.50)

where \( \log L_{M_2} \) denotes the maximal log likelihood of a one-covariate model with upgrade-downgrade effect from the same \( x_{2,t} \) business covariate as in the two-covariate model. Scaled-deviance reduction \( D_{s_{2,1}} \) measures the improvement in going from one-covariate upgrade-downgrade model to a two-covariate model with the same upgrade-downgrade covariate and one more covariate stability-migration.

Similarly, to test the hypothesis in (5.48), we use

\[ D_{s_{2,2}} = \frac{2(\log L_2 - \log L_{M_1})}{\phi}, \]

(5.51)
where \( \log L_{M_1} \) denotes the maximal log likelihood of the one-covariate model with stability-migration effect from the same \( x_{2,t} \) business covariate as in the two-covariate model.

3. Comparison between null model and two-covariate model

Similar to Comparison 1, for two-covariate model, the hypothesis test becomes

\[
H_{30} : \quad \beta_1 = 0 \text{ and } \beta_2 = 0,
\]

\[
H_{31} : \quad \text{not } H_{30}.
\]

And the scaled-deviance difference is

\[
D_{s3} = \frac{2(\log L_2 - \log L_0)}{\phi}.
\]

And \( D_{s3} \) has 2 d.f. \( \chi^2 \) distribution. Note that the alternative hypothesis in test (5.52) can be rewritten as

\[
H_{31} : \quad \text{or } \beta_1 \neq 0 \text{ and } \beta_2 = 0;
\]

\[
\quad \text{or } \beta_1 = 0 \text{ and } \beta_2 \neq 0;
\]

\[
\quad \text{or } \beta_1 \neq 0 \text{ and } \beta_2 \neq 0.
\]

So if any one of the previous tests (5.45) (5.47) (5.48) suggests significance, then we can reject the null hypothesis \( H_{30} \).

5.4 Observed variance of maximum likelihood estimate

We utilize the ML estimation method to obtain the parameter estimates in both one-covariate and two-covariate discrete transition models. In general, under some
regularity conditions, maximum likelihood estimates (MLEs) possess some asymptotic properties. Here, we approximate the true variance of the ML estimate by Fisher observed information matrix.

We rewrite $p_{t,i,j}$ the transition probability from rating $i$ to $j$ in year $t$ as $p_{t,i,j|\theta}$, where $\theta$ is the parameter vector made up by business condition coefficients and baselines $\pi_{i,j}, i \neq j$. $\theta$ is defined for three generalized linear models (GLMs), see (5.29) (5.30) (5.31) for details. So the log likelihood function is

$$\log L = \sum_{t=1}^{19} \sum_{i=1}^{7} \sum_{j=1}^{9} n_{t,i,j} \log p_{t,i,j|\theta}$$

$$= \sum_t \sum_i f_{t,i|\theta};$$

where $f_{t,i|\theta} = \sum_{j} n_{t,i,j} \log p_{t,i,j|\theta}$. Under Assumption 4, with known $x_t$, $f_{t,i|\theta}$ are independent. Because the Multinomial distribution belongs to the exponential family, the interchangeability between the orders of differentiation w.r.t the parameters and integration over the sample space allows us have the following moment identities,

$$E_{\theta} \left( \frac{\partial \log L}{\partial \theta} \right) = 0; \quad (5.56)$$

$$Var_{\theta} \left( \frac{\partial \log L}{\partial \theta} \right) = \left[ -E_{\theta} \left( \frac{\partial^2 \log L}{\partial \theta^2} \right) \right]^{-1}. \quad (5.57)$$

The right side of the above equation (5.57) is the Fisher information matrix. By the independence of $f_{t,i|\theta}$, equations (5.56) and (5.57) become

$$\sum_t \sum_i E_{\theta} \left( \frac{df_{t,i|\theta}}{d\theta} \right) = 0; \quad (5.58)$$

$$Var_{\theta} \left( \frac{d \log L}{d \theta} \right) = \left[ -\sum_t \sum_i E_{\theta} \left( \frac{d^2 f_{t,i|\theta}}{d \theta^2} \right) \right]^{-1}. \quad (5.59)$$
The number of $f_{t,i|\theta}$ values is $T$, the total number of $t$ indices times $I$, the total number of $i$ indices. If the number of $f_{t,i|\theta}$ is large, then the MLEs $\hat{\theta}$ are asymptotically efficient provided that the assumed model is correct and that the derivative is computed at the true parameter point. Here, the total number of $f_{t,i|\theta}$ is $19 \times 7 = 153$. Therefore, we can approximate the true ML variance by the observed Fisher information matrix:

$$V_{ar}(\hat{\theta}) = \left[ -\sum_t \sum_i \left( \frac{\partial^2 f_{t,i|\hat{\theta}}}{\partial \hat{\theta}^2} \right) \right]^{-1}.$$ (5.60)

Over-dispersion exists in this study. Referring to [34], the independence between parameter estimation and over-dispersion allows us to inflate the above estimated covariance matrix by the over-dispersion factor $\phi$. See (5.43). So,

$$\tilde{V}_{ar}(\hat{\theta}) = \phi \left[ -\sum_t \sum_i \left( \frac{\partial^2 f_{t,i|\hat{\theta}}}{\partial \hat{\theta}^2} \right) \right]^{-1}.$$ (5.61)

The over-dispersion suggests that models may be inadequate and the asymptotic normality of ML estimates may be invalid. However, if we assume the dispersion-factor is constant, although this is very hard to verify, then we still could retain the dispersion-adjusted asymptotic normal distribution. This assumption could help us to test the significance of covariates in business-conditioned models.

For a one-covariate model with constant dispersion $\phi$, suppose the first element in variable vector $\theta$ is $\beta$ then the dispersion-adjusted variance, can be estimated as the $(1,1)$ element, of the covariance matrix derived from the Fisher information matrix, say $\sigma^2_{\beta}$, inflated by $\phi$. Moreover, for large sample, we assume

$$\hat{\beta} \sim \mathcal{N} (\beta, \phi \cdot \sigma^2_{\beta}),$$ (5.62)
where $\mathcal{N}(\beta, \phi \cdot \sigma^2_\beta)$ denotes a Normal distribution with mean $\beta$ and variance $\phi \cdot \sigma^2_\beta$. This large sample normality assumption in the presence of over-dispersion has not been theoretically proved.

Therefore, an alternative way to test hypothesis (5.45) is to use a $t$ statistic which is assumed to have Student’s $t$ distribution.

\[
t = \frac{\hat{\beta}}{\sqrt{\hat{\phi} \cdot \hat{\sigma}^2_\beta}}. \tag{5.63}
\]

If the level of significance is set to be $\alpha \times 100\%$ and the $t$ distribution has cutoff value $t_\alpha$, then any statistic in (5.63) greater than $t_\alpha$ suggests a statistically significant non-zero covariate coefficient $\beta$ in the one-covariate model.

For the two-covariate model, similarly, we have,

\[
\hat{\beta}_1 \sim \mathcal{N}(\beta_1, \phi \cdot \sigma^2_\beta_1), \tag{5.64}
\]

and

\[
\hat{\beta}_2 \sim \mathcal{N}(\beta_2, \phi \cdot \sigma^2_\beta_2). \tag{5.65}
\]

To test $H_{11}$ in (5.47), we can use

\[
t_1 = \frac{\hat{\beta}_1}{\sqrt{\hat{\phi} \cdot \hat{\sigma}^2_\beta_1}}. \tag{5.66}
\]

To test $H_{21}$ in (5.48), the $t$ statistic is

\[
t_2 = \frac{\hat{\beta}_2}{\sqrt{\hat{\phi} \cdot \hat{\sigma}^2_\beta_2}}. \tag{5.67}
\]

Any $t_1$ or $t_2$ greater than the $t$ distribution cutoff value will reject the null hypothesis $H_{10}$ or $H_{20}$, respectively.
Chapter 6

Continuous Transition Probability Models

One contribution made by this study is the proposal of a transition probability model over any given continuous time horizon. Almost all of the existing credit rating transition models (Jarrow and Turnbull (1997) [27], JP Morgan&Co.’s CreditMetrics (1997) [2], Nickell et al (2000) [36]) are built on a certain discrete time interval which may or may not coincide with the actual observation time interval – one year. This time framework limitation may result from the discreteness of rating transition observation time. But, in practice, credit rating transition activities can happen at any arbitrary time point. The occurrence of credit rating migration is continuous. Since a change in an issuer’s credit rating status immediately affects the market value of their outstanding debt, a continuous transition model would be more appropriate and useful for modeling credit risk. Moreover, the time factors in pricing credit risk include not only the random and continuous credit change occurrence time but
also the length of the remaining time to maturity. Obviously, the latter time factor is continuously decreasing as the product is continuing alive to maturity. This is another reason for modeling continuous credit rating migrations.

The connection between discrete transition modeling and continuous transition modeling is mainly derived from the assumption of Markovian transition process. If the transition process is Markovian and stationary, then for any time interval, the transition matrix is the result of the accumulating effect of a constant instantaneous transition matrix over that time period. The matrix exponential algorithm is equivalent to a compounding technique. It is applied on an instantaneous transition matrix to obtain the transition probability over any continuous time interval. For a nonstationary transition situation, we have one-covariate and two-covariate instantaneous transition models depending on known time-dependent business time series in a way very similar to the preceding discrete models.

The accomplishment of MLE relies completely on numerical optimization methods because matrix exponential computation dramatically increases the complexity of the maximizing log likelihood function. Thus, it is not possible to find the MLE analytically. Powell’s quadratic optimization was adopted to estimate the MLE instantaneous baselines and covariate coefficients.

Most of our previous notation in discrete modeling is kept in use. $P(t_0, t_1)$ denotes the transition matrix from time $t_0$ to any time $t_1$ with $t_1 \geq t_0$; and the transition probability from rating $i$ to $j$ is denoted as $p_{i,j}(t_0, t_1)$. When $t_0 = 0$, $P(0, t_1)$ and $p_{i,j}(0, t_1)$ become $P(t_1)$ and $p_{i,j}(t_1)$. At any time $t$, the instantaneous transition matrix for the next infinitesimal time period is defined as $\Lambda(t)$. The instantaneous transition
probability from rating $i$ to $j$ is the $(i,j)^{th}$ element, say, $\lambda_{i,j}(t)$.

6.1 Data preparation

6.1.1 Square transition matrix modification

As we are going to show in the following continuous modeling section, for a stationary Markov transition process, a continuous transition probability matrix would be in the exponential form of a constant instantaneous transition matrix. This matrix exponential algorithm requires that the instantaneous transition matrices as well as the estimated continuous transition matrices be square. In this section, we specify how to modify the observed $7 \times 9$ rectangular one-year transition matrices to $9 \times 9$ square matrices.

The observed seven initial ratings are \{AAA, AA, A, BBB, BB, B, CCC\}; and the nine final ratings are \{AAA, AA, A, BBB, BB, B, CCC, D, N.R.\}. The most natural way to square a transition matrix is to extend the initial rating set by adding rating D and N.R. because all observed information would be complete and untouched. The absorbing property of D and N.R, which is a truth for D rating and which is an assumption for N.R. rating, would enable us to extend the transition matrix with certainty: $p_{8,8}(t_1, t_2) = 1$ and $p_{9,9}(t_1, t_2) = 1$ for any $t_1, t_2$.

In continuous modeling, we still keep the same transition distribution assumptions – independent Markovian transition processes on square $9 \times 9$ transition matrices.
6.1.2 Monthly business time series modification

The same 12 economic variables are used in the covariate vector $\mathbf{x}_t$ during 1981 – 1999. Again, they are GDP, PAYRL, UNEMP, CPI, PPI, R3M, R6M, R1Y, GS05, GS10, GS30 and SPREAD. These time series are modified in the same way as in the Discrete Chapter but in *monthly* time unit. For quarterly observed variables GDP, CPI and PPI, we simply use quarterly data as monthly observations. Finally, there are $19 \times 12 = 228$ monthly adjusted observations for each economic covariate.

6.2 Continuous transition probability models

In our study, the distinction between stationary and nonstationary Markovian transition processes is the presence of the transition dependence on time. This time-dependence is characterized by the dependence of covariate vector $\mathbf{x}_t$ on $t$.

6.2.1 Derivation of the constant instantaneous transition matrix

Obviously, time-homogeneity of transitions can lead to stationarity in Markov process. There are two circumstances which enable us to make the time-homogeneous i.e. stationary Markov assumption. One is that the credit transitions are independent from business conditions $\mathbf{x}_t$; while the other is that the business covariate $\mathbf{x}_t$ is constant at any time $t$.

Circumstance 1 of stationary Markov transitions
If we assume that there is no business dependence at all of credit migrations, then any transition rate will be constant for a certain time interval. Therefore, an issuer’s transition activity is only affected by its rating-inherent factor.

Notice this business independence assumption is also the concept of the null continuous model without business condition dependence. We will discuss the null model in the next section.

**Circumstance 2 of stationary Markovian transitions**

If we assume constant business covariates, then the model’s dependence on time through this covariate will disappear.

This assumption is feasible if we constrain our modeling within a very short time interval, say \((t, t + \Delta t)\) and \(\Delta t\) is infinitesimal. Very little change in the business covariate \(x_t\) during this time period is expected.

When the time interval we consider goes to zero, naturally, the probability of an issuer’s rating making a transition to any other ratings goes to zero too. In matrix sense, we have

\[
\lim_{t \to 0^+} P(t) = I. \tag{6.1}
\]

By [31], this assumption is equivalent to the continuity of the stationary transition matrix \(P(t)\). And consequently, any stationary transition rate from rating \(i\) to rating \(j\) is differentiable w.r.t. time \(t\) at \(0^+\). That is, \(\lim_{t \to 0^+} \frac{p_{i,j}(t)}{t}\) is finite.

Under the assumption that credit transitions follow a stationary Markovian process and the above continuity property (6.1) of a transition matrix, the square transition matrix from time zero to any continuous time \(t > 0\) can be expressed as below.
Firstly, under transition Markovian Assumption 3,

\[ P(t) = P(t - \tau)P(t - \tau, \tau) \text{ where } \tau \in (0, t); \quad (6.2) \]

and by stationarity, \( P(t - \tau, \tau) = P(0, \tau) = P(\tau), \) hence

\[ = P(t - \tau)P(\tau). \quad (6.3) \]

Subtracting \( P(t - \tau) \) from both sides of the equation, we have

\[ P(t) - P(t - \tau) = P(t - \tau)[P(\tau) - I]. \quad (6.4) \]

Because \( P(0) = I, \) that is

\[ P(t - \tau)[P(\tau) - P(0)]. \quad (6.5) \]

Divide both sides of equation (6.5) by \( \tau, \)

\[ \frac{P(t) - P(t - \tau)}{\tau} = P(t - \tau) \frac{P(\tau) - P(0)}{\tau}. \quad (6.6) \]

Since \( p_{i,j}(t) \) is differentiable for all \( i, j \) by \( P(t) \) continuity, we take the limit by letting \( \tau \to 0^+ \),

\[ \frac{\partial P(t)}{\partial t} = P(t) \lim_{\tau \to 0^+} \frac{\partial P(\tau)}{\partial \tau}. \quad (6.7) \]

According to Karlin and Taylor (1975) [31], the solution is

\[ P(t) = e^{\Lambda t} \]

\[ = I + \Lambda t + \frac{\Lambda^2 t^2}{2!} + \frac{\Lambda^3 t^3}{3!} + \ldots; \quad (6.8) \]

where matrix \( \Lambda \) is a constant instantaneous transition matrix, i.e. the transition rate matrix for an infinitesimal period of time beginning at time zero. Here, exponential
algorithm (6.8) is equivalent to a compounding mechanism which connects any continuous transition rates with a constant instantaneous transition rate if stationarity holds. So transition matrix $P(t)$ is a continuously compounded result of instantaneous transitions $\Lambda$ over time $(0, t)$. By (6.8),

$$\lim_{\tau \to 0^+} \frac{\partial P(\tau)}{\partial \tau} = \lim_{\tau \to 0^+} \frac{\partial e^{\Lambda \tau}}{\partial \tau},$$

$$= \lim_{\tau \to 0^+} \frac{\partial}{\partial \tau} \left( I + \Lambda \tau + \frac{\Lambda^2 \tau^2}{2!} + \frac{\Lambda^3 \tau^3}{3!} + \ldots \right),$$

because $\Lambda$ is constant,

$$= \lim_{\tau \to 0^+} \Lambda \left( I + \Lambda \tau + \frac{\Lambda^2 \tau^2}{2!} + \frac{\Lambda^3 \tau^3}{3!} + \ldots \right),$$

$$= \Lambda.$$

In summary,

$$\Lambda = \lim_{\tau \to 0^+} \frac{\partial P(\tau)}{\partial \tau}, \quad (6.9)$$

where for any $i, j$,

$$\lambda_{i,j} = \lim_{\tau \to 0^+} \frac{\partial p_{i,j}(\tau)}{\partial \tau}.$$ 

With the unitary row summation constraints in $P(\tau)$,

$$\lambda_{i,i} = \lim_{\tau \to 0^+} \frac{\partial}{\partial \tau} \left( 1 - \sum_{j \neq i} p_{i,j}(\tau) \right).$$

Or the expression is

$$\lambda_{i,j} = \begin{cases} 
\lim_{\tau \to 0^+} \frac{\partial p_{i,j}(\tau)}{\partial \tau} & \text{if } j \neq i, \\
-\sum_{j \neq i} \lambda_{i,j} & \text{if } j = i. 
\end{cases} \quad (6.10)$$

By stationarity, instantaneous transition rates $\lambda_{i,j}(t)$ are constant at any time $t$. 

Therefore, the unitary row summation constraints on $P(t)$ are transformed to the zero row summation constraints on $\Lambda$. In summary, for stationary transitions, the transition probability matrix for any given continuous time period has form (6.8). The matrix $\Lambda$ is the constant instantaneous transition matrix expressed in (6.10).

### 6.2.2 Instantaneous transition probability models

In this section, we will emphasize the modeling of instantaneous transition probabilities. Without business condition independence, the transitions would be stationary and the instantaneous transition matrix would be constant. To verify that the credit rating migration is influenced not only by an individual’s own inherent property but also by current business conditions, we need a model depending on the business covariate vector $x_t$, changing with time $t$. Moreover, this time-dependent covariate results in a non-stationary transition process.

Again, under the Markovian process Assumption 3, a continuous time transition matrix from time zero to $t$ is described as

$$P(t) = \prod_{k=0}^{m-1} P(k\Delta t, (k + 1)\Delta t)$$ (6.11)

where $\Delta t = \frac{t}{m}$. Obviously, $m \to \infty \iff \Delta t \to 0^+$. For large $m$, $\Delta t$ is so small that business conditions can be assumed to give a constant $x \sim_{k\Delta t}$ within each sub-interval $[k\Delta t, (k + 1)\Delta t)$ for any $k = 0, 1, \ldots, m - 1$. This agrees with the stationarity Circumstance 2 defined in the preceding section. Thus, transitions are stationary within these sub-intervals because of infinitesimal $\Delta t$. Over $[k\Delta t, (k + 1)\Delta t)$ define the constant instantaneous transition matrix as
\( \Lambda(k\Delta t) \). Then

\[
P(k\Delta t, (k + 1)\Delta t) = e^{\Lambda(k\Delta t)\Delta t}, \quad (6.12)
\]

The continuous transition matrix (6.11) then is calculated as

\[
P(t) = \prod_{k=0}^{m-1} e^{\Lambda(k\Delta t)\Delta t}; \quad (6.13)
\]

where \( \Lambda(k\Delta t) \) is a constant instantaneous transition matrix for any time within the sub-interval \([k\Delta t, (k + 1)\Delta t)\). For different sub-intervals, \( \Lambda(k\Delta t)\Delta t \) may be different because of the varying business covariates with time. The relationship (6.13) between any continuous transition matrix \( P(t) \) and constant instantaneous transition matrix \( \Lambda(k\Delta t) \) enables us to model \( \Lambda(k\Delta t) \) instead of \( P(t) \).

Similar to discrete modeling, we present the time dependence of instantaneous transition probabilities by the dependence on some known business covariate. The general model of instantaneous transition probability is a function of the baseline transition factor and known business covariates. That is, matrices \( \Lambda(k\Delta t) \) are dependent on the business covariate vector \( x_{k\Delta t} \) for \( k = 0, 1, \ldots, m - 1 \), in a way similar to that in discrete modeling. Specifically,

\[
\lambda_{i,j}(t) = \begin{cases} 
  \pi_{i,j} \exp \sum_{l=1}^{L} (\beta_l x_{l,t} m_{l,i,j}) & \pi_{i,j} \geq 0 \quad \text{if } j \neq i, \\
  -\sum_{j' \neq i} \lambda_{i,j'}(t) & \text{if } j = i.
\end{cases} \quad (6.14)
\]

In the next three subsections, we discuss three specifications of this general transition model (6.14).
6.2.2.1 Null instantaneous transition model

As in the discrete model, a null model must be business-free and with the same framework as the corresponding business-covariate models. The null model is built under the hypothesis,

\[ H_0 : \beta_l = 0; \quad \text{for all} \ l = 1, 2, \ldots, L. \]  

(6.15)

Thus, the null model should have \( L = 0 \) in (6.14), which is equivalent to the stationary Circumstance 1 in previous Section 6.2.1. That is, the business-independence results in a stationary Markovian transition process under null model. The above general model (6.14) is reduced to

\[ \lambda_{i,j}(t) = \begin{cases} 
\pi_{i,j} & \text{if} \ j \neq i, \\
- \sum_{j' \neq i} \pi_{i,j'} & \text{if} \ j = i,
\end{cases} \]  

(6.16)

for any time \( t \).

6.2.2.2 One-covariate instantaneous transition model

Here, we only consider a one covariate in model (6.14). Below \( m_{i,j} \) is the \((i, j)th\) element either in \( M_1 \) (the matrix form of stability-migration effects) or in \( M_2 \) (the matrix form of upgrade-downgrade effects). The Subsection (6.1.1). The known business covariate \( x_t \) can be any one of twelve monthly business time series. Our model for the impact of \( x_t \) on instantaneous rates is

\[ \lambda_{i,j}(t) = \begin{cases} 
\pi_{i,j} \exp(\beta x_t m_{i,j}) & \pi_{i,j} \geq 0 \quad \text{if} \ j \neq i, \\
- \sum_{j' \neq i} \lambda_{i,j'}(t) & \text{if} \ j = i.
\end{cases} \]  

(6.17)
where \( \pi_{i,j} \) is the baseline instantaneous probability of transition from rating \( i \) to \( j \), which is only related to the rating \( i \) inherent transition property. Component \( \exp(\beta x_t m_{i,j}) \) is the business condition adjustment to the baselines. These adjustments could be either relative upgrade-downgrade change or relative stability-migration change.

6.2.2.3 Two-covariate instantaneous transition model

The two-covariate instantaneous transition model is another version of the general instantaneous transition model (6.14) with two covariates. For \( L = 2 \), we have

\[
\lambda_{i,j}(t) = \begin{cases} 
\pi_{i,j} \exp(\beta_1 x_{1,t} m_{1,i,j} + \beta_2 x_{2,t} m_{2,i,j}) & \pi_{i,j} \geq 0 \text{ if } j \neq i, \\
- \sum_{j' \neq i} \lambda_{i,j'}(t) & \text{if } j = i;
\end{cases}
\]  

(6.18)

where \( m_{1,i,j} \) and \( m_{2,i,j} \) are the \((i,j)\)th element of \( M_1 \) and \( M_2 \), respectively. Other parameters have the same interpretations as in the preceding one-covariate model.

Interestingly, if we consider an issuer’s default as a failure event and classify every other rating status as a survival event, then \( \lambda_{i,8}(t) \) is equivalent to the hazard function in survival analysis. This is because

\[
\lambda_{i,8}(t) = \lim_{t \to 0^+} \frac{p_{i,8}(t)}{t};
\]  

(6.19)

by \( p_{i,8}(0) = 0 \),

\[
= \lim_{t \to 0^+} \frac{p_{i,8}(t) - p_{i,8}(0)}{(1 - p_{i,8}(0))t};
\]  

(6.20)

denote \( T_D \) as default time,

\[
= \lim_{t \to 0^+} \frac{Pr(0 \leq T_D \leq t | T_D \geq 0)}{t}.
\]  

(6.21)
Therefore, $\lambda_{i,8}(t)$ is the probability that a time zero $i$-rated ($i \neq 8$) issuer defaults in the infinitesimal interval $(t, t + \Delta t)$.

For the null model, the hazard rate is constant for each of seven non-default initial ratings. For one-covariate and two-covariate models, (6.17) and (6.18) are exactly Cox regression models with constant baselines $\pi_{i,j}$.

### 6.3 Maximum likelihood estimation and optimization implementation

#### 6.3.1 Likelihood function

For any one year period $[T, T + 1)$, we equally partition it into $m$ sub-intervals. Therefore, for each sub-interval, the length is $\Delta t = \frac{1}{m}$. These sub-intervals are: $[T, T + \Delta t), [T + \Delta t, T + 2\Delta t), \ldots, [T + (m - 1)\Delta t, T + m\Delta t)$. As integer $m$ goes to infinity, the time length $\Delta t$ is positive infinitesimal, hence we assume that the rating transition within each sub-interval has constant instantaneous transition rate matrix. Also by the Markovian property, the one-year transition matrix can be written as

$$P(T, T + 1) = \prod_{k=0}^{m-1} e^{A(T + k\Delta t)\Delta t}.$$  \hfill (6.22)

In the Discrete Chapter, the log likelihood function for 19 one-year transitions was derived as

$$\log L = \sum_{T=1}^{19} \sum_{i=1}^{7} \sum_{j=1}^{9} n_{T,i,j} p_{i,j}(T, T + 1);$$  \hfill (6.23)

where $p_{i,j}(T, T + 1)$ is the $(i, j)$th element of matrix $P(T, T + 1)$. Obviously, for the null model, unknown parameters include $\pi_{i,j}$ for $i = 1, 2, \ldots, 7; j = 1, 2, \ldots, 9$; for
a one-covariate instantaneous transition model, unknown parameters are $\beta$ and $\pi_{i,j}$ for $i = 1, 2, \ldots, 7; j = 1, 2, \ldots 9$; for a two-covariate instantaneous transition model, unknown parameters include $\beta_1, \beta_2$ and $\pi_{i,j}$ for $i = 1, 2, \ldots, 7; j = 1, 2, \ldots 9$.

We estimate these parameters in three instantaneous transition models by maximizing the log likelihood function (6.23) over the period 1981 – 1999. With the modified monthly data and relation (6.22), we calculate the one-year transition matrix as the product of twelve ($m = 12$) matrix exponentials. The difficulty is that the log likelihood function is too complicated to be solved analytically. The complexity of $p_{i,j}(T, T+1)$ is caused by the great number of matrix products and matrix summations in the matrix exponential algorithm (6.22). Consequently, analytically maximizing the log likelihood function (6.23) is impossible for us. Therefore, we turn to seek a solution with a numerical high-dimensional optimization method – Powell’s quadratic optimization.

The number of baseline parameters $\pi_{i,j}$ for $i = 1, 2, \ldots, 7; j = 1, 2, \ldots 9$ with $i \neq j$ is 56. Because a high dimension in parameter space can lead to great difficulty in optimizing, we first need to reduce the number of parameters.

For any very short time period, the probability of an issuer’s rating transiting to any other ratings is very low. From 19 one-year transition observations, other than transition rates to D and N.R., the off-diagonal rate diminishes as the distance between final rating and initial rating increases. So, we restrict our interest to estimate the within-two-step instantaneous transition probabilities. In other words, we do not consider any transition (excluding D. and N.R. transitions) beyond two-step upgrade or two-step downgrade; we assume those transitions have zero likelihood in an in-
finitesimal time period. In matrix language, only penta-diagonals in the $\Lambda$ matrix are nontrivial for us to project instantaneous transition probabilities. That is, for $i, j = 1, 2, \ldots, 7$, $\pi_{i,j} = 0$ if $|i - j| > 2$.

Historically, there are almost no rating AAA, AA and A defaults. The best estimates of corresponding instantaneous rates are believed to be very close to zero. As for other ratings' transitions to D, the observed default rate increases as the initial rating deteriorates. Observed average default rates for ratings BBB, BB, B, and CCC are not negligible – starting with an average default rate for BBB 0.2%, they move on to CCC with about 21%. So we set $\pi_{i,8} = 0$ for $i = 1, 2, 3$.

Finally, the number of $\pi_{i,j}$s is reduced to 30. We notice that the compounding effect of the exponential algorithm can still generate positive transition probabilities $p_{i,j}(t)$ where $\pi_{i,j} = 0$.

### 6.3.2 Powell’s quadratic numerical optimization

The parameter vectors in instantaneous transition models have high dimensions. A modified Powell’s quadratic optimization method is adopted to maximize likelihood (6.23). The program code can be found at ‘Numerical Recipe’ web site.

Because of the complexity of the likelihood function, there is no guarantee of a unique maximum. To numerically check the maximum likelihood estimate’s (MLE) uniqueness, we need to start with different starting vectors to check whether they converge to the same point.

The maximization is only derived from the observed transitions. That is, tran-
itions from initial ratings $i \in \{\text{AAA, AA, A, BBB, BB, B, CCC}\}$ to final ratings $j \in \{\text{AAA, AA, A, BBB, BB, B, CCC, D N.R.}\}$. The last two rows of the estimated one-year transition matrix $P(T, T+1)$ are not used by optimization due to the absorbing property of rating D and N.R.. Thus, their transition rates are known without being observed.

6.4 Goodness-of-fit test, over-dispersion and scaled-deviance

We adopted Pearson’s $\chi^2$ statistic, again, to measure the goodness-of-fit of our continuous model. We need to calculate the degrees of freedom for $\chi^2$ to be consistent with the transition matrix adjustment and instantaneous transition probability reduction.

There are $19 \times 7 \times 9$ different transitions in total and $19 \times 7$ zero summation constraints. Because we focus on the modeling of penta-diagonal instantaneous transition probabilities, the last four instantaneous default probabilities and seven N.R. instantaneous transition probabilities, the actual number of instantaneous baseline parameters would be 33. Finally the d.f. of $\chi^2$ is $19 \times 7 \times 9 - 19 \times 7 - 33 = 1031$ for the null model, 1030 for the one-covariate model and 1029 for the two-covariate model.

Similarly, we will see in our results that over-dispersion does exist and is greater than that in discrete models. We estimate the over-dispersion factor $\phi$ same as (5.43) in Chapter 5.
The null model, one-covariate model and two-covariate model are nested. The reduction in scaled-deviance from the null model to the one-covariate model can be used to identify a significant business covariate that has impact on the instantaneous transition activities; while the reduction in deviance from the one-covariate to the two-covariate model can be used to investigate how significant the improvement is by adding one more business effect to the one-covariate model. All the difference in scaled-deviance need to be adjusted to scaled-deviance reductions by the dispersion estimate, the same as (5.44).
Chapter 7

Part II Results

In the preceding Chapter 5 and Chapter 6, we proposed a model of a credit rating transition matrix in discrete time and a model of an instantaneous transition matrix in continuous time. These models represented both the credit rating internal transition behavior and the external impact of economic conditions. Two major economic effects on the transition matrix were examined: upgrade-downgrade effect and stability-migration effect. In this chapter, we include the MLE (Maximum Likelihood Estimation) results for the discrete transition probability model and for the continuous transition probability model.

To test the significance of business impact on transition probabilities and hence to identify the most significant business time series, all the results are presented in the nesting order of null model, one-covariate model and two-covariate model.
7.1 Results for discrete transition models

The structure of a discrete model enables us to investigate different initial rating transitions separately. There is a widely held belief that issuers belonging to different categories, investment-grade and speculative-grade, are more unlike than those belonging to same category. We therefore model the discrete transition behavior for complete-grade (including all seven initial ratings), investment-grade (including only initial ratings AAA, AA, A, BBB) and speculative-grade (including only initial ratings BB, B, CCC).

In this section, we show and investigate the discrete MLE results for the complete transition matrix, for an investment-grade transition matrix and for a speculative-grade transition matrix.

7.1.1 Complete-grade transitions

7.1.1.1 Null model

As discussed before, a discrete null model was needed to detect the significance of business condition impact on credit migration. It was required to be business-free such that the comparison of any business-covariate model to this null model is meaningful. With the independent multinomial distribution assumption, under the null model, the estimated constant one-year transition matrix given in equation (5.34) is shown in Table 7.1.

The estimated one-year transition matrix is exactly the same as the average one-year transition rates given in the S&P report. So we understand that S&P uses the
same multinomial assumption to obtain the average transition probabilities. If we rewrite the estimated one-year transition probability from rating $i$ to $j$ in expression (5.34) as

$$\hat{\pi}_{t,i,j} = \frac{n_{t,i,j}}{n_{t,i}}, = \frac{1}{n_{t,i}} \sum_{t=1}^{19} n_{t,i,j},$$

then we find that the number of issuers with rating transition $i$ to $j$ in year $t$ is proportional to the weight of the year $t$ contribution to the estimate. Because the numbers of issuers for recent years are much greater than for early years, this estimated one-year transition matrix tends to reflect more recent transition activities. This estimated rating transition matrix under the null model could be thought of as the pure rating inherent transition matrix which evens out all economic effects.

The following paragraph summarizes these rating intrinsic properties observed from Table 7.1.

First, higher ratings had lower default probability. Historically, AAA had never had any defaults; AA and A were almost zero default-risky; the default probabilities jumped dramatically from the BBB-rated 0.2% to CCC-rated 21%. Secondly, the rating stability decreased as the credit rating status lowered. These diagonal elements, which stand for the stable probabilities for each initial rating, showed that investment-grades would retain the same rating in a year with above 80% probability while speculative-grades would have lower stable probabilities around 53% to 75%. Finally, regardless of the transitions to D and N.R., rating migration probabilities were diminishing with the increasing transition distance. Moreover, rating migrations were more likely to happen within two steps in both upgrade and downgrade directions for one year time horizon. We inferred that any transition exceeding two
steps would be unlikely if the transition time period is much shorter than one year.

Along with the estimation of the null transition matrix, the maximum log likelihood and goodness-of-fit statistic $\chi^2$ were also calculated. The maximized null log likelihood function had value $-31732.87$; goodness-of-fit $\chi^2$ statistic was 2786.14 with 1013 d.f.; thus the estimated over-dispersion factor was $2.75 = 2786.14/1013$. Comparing to a $\chi^2$ 95% cutoff with 1013 d.f., 1088.16, the model was under-fitting. The greater than one dispersion factor estimate, 2.75, suggested the existence of a random over-dispersion effect which could not be captured by this null model.

7.1.1.2 The one-covariate model

The one-covariate discrete model (5.14) adjusted the rating-intrinsic baseline transition probabilities $\pi_{i,j}$ with an extrinsic transition impact from business conditions. Two major impacts defined were upgrade-downgrade effect and stability-migration effect, through which business conditions might affect overall rating migrations. Specifically, we try to understand how different business condition covariates affected the relative change between the transition probabilities of upgrades and downgrades; and the relative change between the transition probabilities of rating stability and rating migration.

The matrix representation of stability-migration was the $M_1$ matrix and that of upgrade-downgrade effect was the $M_2$ matrix and as defined in Section 5.1. The rating-inherent transition probabilities were quantified through baseline transition probabilities $\pi_{i,j}$. Recall that the business effect was modeled as a product adjustment on baselines in the form of $\exp(\beta x_i m_{i,j})$. We estimated baselines $\pi_{i,j}$ and a
business covariate $\beta$ coefficient by maximizing the log likelihood function, which was fulfilled by an optimization algorithm GCDO (Generalized Cyclic Descent Optimization).

One major goal in constructing these two nesting models, null and one-covariate, was to test whether a business covariate had significant impact on the transition matrix. We tested this by two methods. One was using the $\chi^2$ distributed scaled deviance deduction discussed in Section 5.3 and the other was using the asymptotic normality of the covariate coefficient estimate discussed in Section 5.4.

For the scaled-deviance $\chi^2$ test, the difference in number of parameters of the null model and one-covariate model is one. So we need to compare the estimated scaled-deviance with 1 d.f. $\chi^2$ 95% cutoff, 3.84. Any scaled-deviance $D_s_1$ greater than 3.84 indicates significant stability-migration impact from business covariate; similarly, any Any greater than 3.84 scaled-deviance $D_s_2$ indicates significant upgrade-downgrade impact from the business covariate.

As for the asymptotic $t$ test on $\hat{\beta}$, which was expressed in (5.63), we compare the Student’s $t$ statistics with the two-sided $t$ distribution 95% cutoff, 1.96, to test the significance of the business covariate impact.

Table 7.2 and Table 7.3 have the results for a discrete one-covariate model with stability-migration impact and for that with upgrade-downgrade impact, respectively. They include the reduction of scaled-deviance from the null model $D_{s1}$, covariate coefficient estimate $\hat{\beta}$, standard deviation of $\hat{\beta}$ sd expressed in units of $10^{-2}$, the asymptotic $t$ statistics, and the estimate of the over-dispersion parameter $\hat{\phi}$.

**Stability-migration effect**
Table 7.2 contains the results for the one-covariate model with business stability-migration transition effect.

According to test (5.45), any reduction of scaled-deviance $D_{s1}$ greater than 3.84 (95% cutoff of 1 d.f. $\chi^2$ distribution) indicates significance of business impact on the relative upgrade-downgrade transition probabilities. From Table 7.2, only CPI and PAYRL were not significant variables because their $D_{s1}$ values were less than 3.84.

An alternative way to test the significance of business impact is to use the asymptotic $t$ test detailed in Section 5.4. We found the $t$ test result was consistent with the $D_{s1}$ $\chi^2$ test. Again, CPI and PAYRL have $t$ statistics smaller than the cutoff 1.96.

SPREAD was the most important covariate affecting stability-migration transitions for it showed the greatest improvement from the null model. It had $D_{s1} = 16.36$. All the interest rates give significant reduction in scaled-deviance from the null model. However, the relation between the significance of impact and the rate time span was not clear. Three-month interest rates and 10-year interest rates had greater maximum likelihood improvement than the the other interest rates. In general, all interest rates outperformed significant covariates PPI, UNEMP and non-significant covariates CPI and PAYRL.

Recall the discrete one-covariate model,

$$\log p_{t,i,j} = \log \pi_{i,j} + \beta x_t m_{i,j}, \quad (7.1)$$

with restriction that $\sum_{j=1}^{9} p_{t,i,j} = 1$. For testing the stability-migration business effect, we set elements in $\mathbf{M}_1$ as $m_{i,i} = +1$ and rating migration as $m_{i,j} = -1, j \neq i$.

The covariate coefficient $\beta$ in the one-covariate model is proportional to the impact of a one-unit change in the business variable $x_t$ on the log of transition probabilities.
Only GDP and PAYRL have positive estimates $\hat{\beta}$ and the other variables have negative estimates. The signs of $\hat{\beta}$ coincided with common expectation and observation. That is, a positive change in GDP, PAYRL and negative change in CPI, PPI, UNEMP, rates indicate a positive growth of the economy; while negative change in GDP, PAYRL and positive change in CPI, PPI, UNEMP, rates indicate a decline in the economy. So when the economy is in expansion or recovery (positive growth), there are relatively less issuers’ ratings being changed to other ratings ($m_{i,j} = -1, j \neq i$) and more of them remaining same ($m_{i,i} = +1$); when the economy is in recession or slowdown (negative growth), there are relatively more issuers’ ratings migrating to other ratings and less remaining the same. This was also founded in our Part I study – business cycles and credit cycles are correlated.

For example, the $\hat{\beta}$ for SPREAD is -0.10. So when the change in yield difference between the almost riskless AAA-rated bond and risky BBB-rated bond is one-unit positive, which implies an increase in credit risk of the BBB-rated bonds, all issuers will be less likely (-0.10 change in $\log p_{t,i,i}$) to stay the same at their initial rating and more likely (0.10 change in $\log p_{t,i,j}, j \neq i$) to change their credit quality.

Compared to the null model over-dispersion parameter estimate $\hat{\phi} = 2.72$, significant one-covariate stability-migration models fit the data better with less over-dispersion of estimates. That shows that incorporating business impacts can partially reduce the over-dispersion of the null model.

**Upgrade-downgrade effect**

From Table 7.3, all 12 business time series had significant impact on upgrades and downgrades. This conclusion was drawn because all $D_{s1}$ were greater than the
χ^2_{1, d.f} 95% cutoff, 3.84. This statement is supported by the asymptotic t test: all large t statistics suggested nonzero covariate coefficients. At the same time, all the one-covariate upgrade-downgrade models fit the data better than a market-free null, because all the over-dispersion factors were less than the over-dispersion factor 2.72 in that null model.

As we mentioned, D_{si} is the likelihood improvement measure comparing the one-covariate model to null model. By examining D_{si} for all the 12 business variables, we found that PAYRL (payroll) was the most important economic indicator to affect the relative transition probabilities between upgrades and downgrades. The second was GDP and the third was SPREAD.

The discrete one-covariate model with upgrade-downgrade effect is

\[ \log p_{t,i,j} = \log \pi_{i,j} + \beta x_t m_{i,j}, \]  

(7.2)

with the restriction that \( \sum_{j=1}^{9} p_{t,i,j} = 1 \). Upgrades in M_2 were assigned \( m_{i,j} = +1, j < i \) and downgrades were assigned \( m_{i,j} = -1, j > i \).

So when the economy is in expansion or recovery (positive growth), there are relatively less downgrades and more upgrades; when the economy is in recession or slowdown (negative growth), there are relatively more downgrades and less upgrades. For instance, when the growth rate of GDP is positive, productivity increases, corporate issuers can generate more profit and thus it is more likely for their credit ratings to get upgraded than to get downgraded. Similar interpretations can be made for other business covariates.

Note that shorter interest rates had more significant impact on upgrade-downgrade transitions than longer interest rates. As time spans are increasing from three months
to 30 years., the improvements in scaled deviance are decreasing. The yield difference between Moody’s BBB-rated bond and AAA-rated bond, SPREAD, was most correlated with the credit risk of ratings and was the most important rate covariate to model the upgrade-downgrade transition impact.

We also estimated a baseline transition matrix for one-covariate models. In fact, for each time series and each one-covariate model, the estimates of baseline transition probabilities, \( \hat{\pi}_{i,j} \), were very close to the estimated null transition matrix in Table ???. This fact along with the small covariate coefficient estimates \( \hat{\beta} \) suggest that rating-intrinsic transition behavior is dominant and business condition impacts induce fluctuations in the transition matrix around the business-free baseline transition matrix.

For all 12 covariate time series, the reduction in scaled-deviance in the one-covariate upgrade-downgrade model was much greater than that in the one-covariate stability-migration model. This indicated that business conditions affected the discrete one-year transition by adjusting relative transition probabilities between upgrades and downgrades on inherent rating transition activities.

### 7.1.1.3 Two-covariate model

Table 7.4 contains the ten best two-covariate models in terms of the improvement in scaled-deviance comparing the two-covariate model to the null model. This was denoted by \( D_{s3} \), formulated in (5.49).

\[
D_{s3} = \frac{2(\log L_2 - \log L_0)}{\hat{\phi}};
\]
where $\hat{\phi} = \chi^2/1011$ in a two-covariate model. Since the difference in number of parameter between the null model and the two-covariate model was two, the distribution of $D_{s3}$ was approximately $\chi^2_2$ distributed. The statistic was used to test hypothesis in (5.52).

From the hypotheses $H_{30}$ and $H_{31}$ in (5.54), we know that the test (5.48) is a sub-test of test (5.52). So the significance of a one-covariate model automatically leads to the significance of the corresponding two-covariate model. It was known that all the significance of one-covariate model with the upgrade-downgrade effect. Hence, we can conclude that all the two-covariate model have significant improvement from the null model.

We wanted to test the necessity of extending the corresponding one-covariate models to two-covariate. This is equivalent to testing the hypothesis in (5.47) and (5.48). Let $D_{s1}$ be the likelihood improvement from one-covariate upgrade-downgrade model to a two-covariate model with same upgrade-downgrade covariate, and one more covariate with stability-migration effect; let $D_{s2}$ measure the improvement from one-covariate stability-migration model to a two-covariate model with same upgrade-downgrade covariate and one more covariate with stability-migration effect. See expressions (5.50) and (5.51) for details. Any $D_{s1}$ or $D_{s2}$ greater than 95% $1 \text{ d.f. } \chi^2$ cutoff 3.84 indicates the significant likelihood improvement of adding one covariate with different impact on the corresponding one-covariate model.

An alternative test was the asymptotic t test for estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ with dispersion-inflated standard deviation. See Section 5.4 for details.

In Table 7.4, the ten best two-covariate models are significantly better than the
null because all the $D_{s3}$ are greater than 5.99. The best two-covariate model was SPREAD having stability-migration impact and PAYRL having upgrade-downgrade impact. Both covariates were significant implied by large $D_{s21}, D_{s22}, t_1$ and $t_2$. This optimal two-covariate result was coincident with those in the one-covariate model: PAYRL was the most important covariate M1 in the one-covariate model and SPREAD had the most improvement in the M2 one-covariate model.

PAYRL is the dominant significant covariate in two-covariate model with upgrade-downgrade transition impact. Eight out of ten best models are having PAYRL as upgrade-downgrade covariate.

By checking the scaled-deviance improvements, $D_{s21}$ and $D_{s22}$, from corresponding one-covariate models, we found that the upgrade-downgrade effect is significant for all ten two-covariate models. This is because $D_{s22} > 3.84$. However not all the covariates with the stability-migration effect are significant. The seventh, eighth and ninth models can be reduced to one-covariate models with PAYRL upgrade-downgrade impact because of their non-significant improvement, $D_{s21} < 3.84$. All of the above significance results are also supported by asymptotic $t$ tests, $t_1$ and $t_2$.

When $D_{s22}$ was greater than $D_{s21}$ for all the two-covariate results, this indicated that there was more improvement made by adding an upgrade-downgrade impact to a stability-migration one-covariate model than from adding a stability-migration impact from upgrade-downgrade one-covariate model. This fact again demonstrated the importance of business upgrade-downgrade effect on the transition matrix rather than rating stability in a one-year period.

Smaller over-dispersion parameter estimates than those found for one-covariate
models suggested that adding one more business impact does capture part of the variation unexplained by the one-covariate model.

In summary, the best discrete one-year transition model was the two-covariate model with PAYRL having the most significant upgrade-downgrade effect and SPREAD having the most significant stability-migration effect on overall rating transitions. Business condition tended to affect transition matrix by having impact on relative upgrade-downgrade transition probabilities rather than on relative stability-migration transition probabilities.

### 7.1.2 Investment-grade transitions

The result for complete-grade transition models was derived from jointly modeling and estimating all seven initial rating transition behaviors. In practice, financial practitioners and researchers believe that rating transitions for investment-graded issuers and for speculative-graded issuers are different. For instance, speculative-grade rating transitions are more sensitive to business conditions than investment-grade rating transitions.

For the same discrete transition probability model, if we allow the parameters to be estimated differently for these two rating categories, then we can investigate investment rating transition probabilities and speculative rating transition probabilities separately. By comparing their maximum likelihood estimates (MLEs), we can also detect how business conditions affect their rating transitions differently.

More specifically, for investment-grade and speculative-grade transition probabilities, we have the same null model, one-covariate model and two-covariate model,
proposed as (5.13), (5.14), and (5.16) in Section 5.1. To estimate parameters independently for investment-grade transitions and speculative-grade transitions, the log likelihood function is

\[
\log L = \sum_{t=1}^{19} \sum_{i=1}^{7} \sum_{j=1}^{9} n_{t,i,j} \log p_{t,i,j}
\]

That is, we can independently maximize two separate log likelihood functions: investment-grade log likelihood function \( \log L_{INV} \) and speculative-grade log likelihood function \( \log L_{SPE} \).

This subsection and the next contain the results for investment-grade transitions and speculative-grade transitions, respectively. All the discussion will be similar to those in the combined transition model results.

### 7.1.2.1 Null model

For the null model estimate equation (5.34), the maximized log likelihood was \(-18851.22\); \( \chi^2 \) was 1619.23 with \( d.f. = 579 \); and the over-dispersion factor was 2.80. The 4 estimated investment-grade null transition matrix was the same as the first four rows of Table 7.1.

### 7.1.2.2 One-covariate model

**Stability-migration effect**
Table 7.5 has the result from a one-covariate stability-migration model on investment-grade transition probabilities for 12 business variables. By both reduction of scaled-deviance $D_{s1}$ statistics and $t$ statistics, GDP, CPI, PAYRL and UNEMP are not significant in their effect on investment-grade relative transition probabilities between stability-migration transitions. Among the other significant business variables, SPREAD is the most important with stability-migration effect.

Interestingly, the longer the interest rates, the more significant the stability-migration effect for investment-grade transitions. This was proved by increasing $D_{s1}$ with the increase of time spans of rates. This result may indicate that investment-grade transitions are more affected by long-term interest rate risks than short ones.

The negative estimates of significant covariate coefficient again agree with our expectation: for investment-graded issuers, their ratings more likely stay same under a friendly economic condition (lowering interest rates, reducing inflation indicator CPI) and they are more likely changed to other ratings under a severe economic condition (increasing interest rates, higher inflation indicator CPI).

The estimates of over-dispersion parameter are less than null model estimate, 2.80. So, again, incorporating business condition factor into null model reduced model dispersion.

**Upgrade-downgrade effect**

The upgrade-downgrade one-covariate model result for investment-grade transitions are showed in Table 7.6.

Obviously, all the business covariate have significant effect on investment-grade relative upgrade-downgrade transition probabilities. SPREAD is the most important
variable for this model. Interest rates show same trend as in one-covariate stability-migration model: longer interest rates are more important in affecting investment-grade upgrade-downgrade transitions.

Positiveness of GDP and PAYRL coefficient estimates and negativeness of the other variable coefficient estimates made same economic sense of our model: good economic condition tends to increase investment-grade upgrade transition probabilities and decrease downgrade transition probabilities.

When we compared the reduction of scaled-deviance with one-covariate stability-migration model, we found that all the business covariate were significant comparing to null; That is, the upgrade-downgrade dominant impact over stability-migration effect on transitions remained true for investment-grade.

7.1.2.3 Two-covariate model

Table 7.7 includes the top ten two-covariate models for investment-grade transitions.

All the ten two-covariate models significantly improved the likelihood from investment-grade null model, from one-covariate stability-migration model, and from one-covariate upgrade-downgrade models. The importance of upgrade-downgrade impact over stability-migration impact is not as certain as in complete-grade model result. For two-covariate models, SPREAD was crucial to the stability-migration effect on investment-grade transitions. The two-covariate model with SPREAD having stability-migration effect and GS30 having upgrade-downgrade effect was the best.

In summary, for one-year investment-grade transitions, business condition does
have impact on their transition probabilities; long-term business condition has more impact than short-term business condition; business condition more likely to affect the relative change in transition probabilities between upgrade-downgrade less likely to affect the relative change in transition probabilities between stability-migration.
7.1.3 Speculative-grade transitions

7.1.3.1 Null model

Ratings BB, B, and CCC make up of the category of speculative-grade rating. Similar to investment-grade discussion, the speculative null model one-year transition matrix was same as the last three rows of complete-grade null model one-year transition matrix showed in Table 7.1. The maximized log likelihood function had value $-12881.65$; goodness-of-fit statistic $\chi^2$ was $1166.91$ with $434$ d.f.; and over-dispersion factor was estimated as $2.69$.

7.1.3.2 One-covariate model

Table 7.8 and Table 7.9 contain the results for speculative one-covariate models with stability-migration and with upgrade-downgrade and effect, respectively.

Stability-migration effect

For the investigation of business stability-migration effect on investment-grade rating transitions, both scaled-deviance and $t$ statistics showed that only PAYRL and GDP were significant business variables in Table 7.8. Both covariates had positive $\hat{\beta}$. And GDP was the most important covariate. Interestingly, in investment-grade transition study, PAYRL and GDP were not significant. The difference may be interpreted as that speculative-graded issuers’ rating transitions are more sensitive to economic growth rate than investment-graded issuers’ rating transitions.

Upgrade-downgrade effect
For the investigation of business upgrade-downgrade effect on speculative-grade rating transitions, GDP, PAYRL, short rates (R3M, R6M, R1Y) and SPREAD had significant upgrade-downgrade impact on one-year transition matrix. Among those variables, PAYRL was the most important, SPREAD was the second. The signs of $\hat{\beta}$ for significant covariates still reflected the expected business condition impact, i.e. more upgrade rating transitions and less downgrade rating transitions under friendly economic condition; more downgrade and less upgrades under sever economic condition for speculative-graded issuers.

Comparing the improvement of likelihood from null model, upgrade-downgrade one-covariate model had greater $D_s$, than stability-migration one-covariate model, except for variables UNEMP, which is not significant for both models. So upgrade-downgrade effect is business condition’s prior impact for speculative-grade transitions.

Different from investment-grade transitions, for speculative-grade transitions, the shorter the rate, the more significant the impact on upgrade-downgrade transitions. This fact indicates that speculative-graded rating transitions are more sensitive to the stage of business cycle and have more short-term market-related risk.

Note that unlike the results for investment-grade and complete-grade transition models, SPREAD was not a significant business covariate for speculative-grade by affecting rating stability-migration transitions. This may be because SPREAD is a measure of investment-grade credit risk defined as the yield difference of BBB-rated bond, belonging to investment-grade, to AAA-rated bond.

Above comparisons between investment-grade and speculative-grade demonstrated the impacts on their transitions are different by business condition.
7.1.3.3 Two-covariate model

In Table 7.10, the best two-covariate model was GDP with stability-migration effect and SPREAD with upgrade-downgrade effect. GDP was the dominate variable having significant stability-migration effect in two-covariate model. This was consistent with the result in one-covariate stability-migration model for speculative-grade transitions. Short rates outperformed long rates again with upgrade-downgrade effect on speculative-grade transitions.

Although all ten two-covariate models were significantly fitting better than null models for $D_{s_3} > 5.99$, most models could be reduced to their corresponding one-covariate model. For instance, two-covariate model 5 with GDP+R1Y had non-significant R1Y upgrade-downgrade effect, then it was not significantly different from one-covariate GDP stability-migration model. Similar discussion can be applied on model 7, 8, 9, 10. Finally, the significant two-covariate models are model 1, 2, 3, 4, and 6.

7.2 Results for continuous transition models

To model continuous rating transition probabilities, square instantaneous transition matrix was required. We assumed N.R. to be an absorbing rating. In instantaneous transition matrix, only penta-diagonals, default probabilities of BBB, BB, B, CCC were modeled.

Note that the average transition rate from CCC to AAA is not close zero. This observation was an outlier which created more over-dispersion of our models. This
is because the exponential compounding mechanism of penta-diagonal instantaneous transition matrix had very small fitted transition probability for long-distance transitions. The discrepancy between large observed CCC→AAA rates and small fitted CCC→AAA probabilities increased goodness-of-fit measurement, χ². So we replaced all the observed CCC→AAA transition rates by zero.

### 7.2.1 Null model

Table 7.11 was the MLE estimated instantaneous transition matrix \( \hat{\Lambda}_B \) under null model. Thirty-three baseline parameters included penta-diagonals (excluding diagonals), default probabilities of BBB, BB, B and CCC, and seven N.R. transition probabilities. This instantaneous transition matrix was measured at one-month time horizon.

The one-year transition market-free transition matrix could be approximated by \( e^{12\hat{\Lambda}_B} \). Table 7.12 is the approximated continuous null one-year transition matrix. Comparing to discrete one-year null model transition matrix in Table 7.1, continuously approximated one-year transition matrix had greater probabilities for modeled 33 transition and smaller probabilities for other transitions. This was the result from matrix exponential algorithm.

The continuous null model had a maximized log likelihood function of -31976.80; dispersion factor was 9.76 derived from \( \chi^2 \) 10062.85 with 1031 d.f.. Comparing to discrete null model result (maximum log likelihood=-31732.87; \( \chi^2 \)=2786.14 with 1013 d.f. and \( \phi = 2.75 \)), continuous null model was much more dispersed for its larger \( \chi^2 \) and dispersion factor estimate. Besides the existence of random effect, the less
number of parameters and the exponential smoothness also could lead to large discrepency of model fitted values and observations. Another contribution to this larger over-dispersion than discrete model might be the lack of real instantaneous transition observations. In fact, instantaneous transition probabilities were estimated by approximating one-year discrete transition. This approximation can result in more model dispersion.

Next, we exam how business covariates affect instantaneous transition probabilities with one or two effects.

7.2.2 One-covariate model

Stability-migration effect

Table 7.13 shows the result for continuous one-covariate model with business stability-migration effect.

From the result of $D_{s_1}$, only SPREAD and long interest rates R1Y, GS10 and GS30 are significant; SPREAD was the most important business covariates having stability-migration effect.

The negative estimates of covariate coefficients for SPREAD, R1Y, GS10 and GS30 suggest that when long interest rates the yield between BBB-rated bond and AAA-rated bond increase, the instantaneous upgrade transition probabilities tend to decrease and downgrade transition probabilities tend to increase.

The estimates over-dispersion factor were not obviously reduced from null model’s 9.76. So incorporating one-covariate stability-migration effect into null model did not explain the over-dispersion in null model.
Upgrade-downgrades effect

The result of reduction in scaled-deviance showed in Table 7.14 indicates UNEMP and PPI were not significant; SPREAD and PAYRL were the most important business covariates having instantaneous upgrade-downgrade effect. Only GDP and PAYRL had positive $\hat{\beta}$.

The comparison between $D_{s1}$ in two one-covariate models with same business covariate tells that upgrade-downgrade effect is business condition’s prior impact on instantaneous transition probabilities to stability-migration effect.

Above results are similar to what we found in discrete one-covariate model result. But for this instantaneous one-covariate model, longer interest rates outperformed shorter interest rates. This is different from discrete one-covariate model, where shorter interest rates had more significant one-year upgrade-downgrade effect than longer rates.

7.2.3 Two-covariate model

The top ten best continuous two-covariate result is contained in Table 7.15. All two-covariate models significantly improved likelihood from null model by $D_{s3} > 5.99$; both business covariate with different effects in these 10 models almost equally contributed to the improvement and they were all significant. Same as in discrete two-covariate model, two-covariate model with SPREAD stability-migration effect and PAYRL upgrade-downgrade effect was the optimal continuous instantaneous two-covariate model. SPREAD was a critical business variable for both business effects in two-covariate model, which was consistent with the result in single matrix model.


7.3 Conclusion

The two business condition impacts are defined as upgrade-downgrade effect and stability-migration effect. In last section, we showed results of our generalized linear model of business condition impact on transition probabilities for one-year complete-grade transitions, one-year investment-grade transitions, one-year speculative-grade transitions, and complete-grade instantaneous transitions. Below is the summary of same conclusion we drew from these results.

- Business condition does have significant upgrade-downgrade effect and stability-migration effect on credit rating transition probabilities; and the upgrade-downgrade effect is more significant than the stability-migration effect.

- Small covariate coefficient estimates $\hat{\beta}$ suggest dominant inherent transition behavior and minor but significant transition adjustment from business condition.

- All maximum likelihood results give positive $\hat{\beta}$ for significant variables GDP, PAYRL and negative $\hat{\beta}$ for other significant variables CPI, PPI, UNEMP, r3M, R6M, R1Y, GS05, GS10, GS30 and SPREAD. So, for positive growth rate of productivity, the positive growth rate of new non-farm payroll created, fewer inflation signs, lower unemployment rate, smaller short and long interest rates, narrower yield spread between Moody’s BBB-rated bond and AAA-rated bond, it is more likely for issuers to remain their credit rating same, and for those with rating transition, it is more likely for them to have upgraded transitions.

Besides above similarity in results for different models, there are some differences between discrete transition model and continuous transition, between investment-
grade one-year transition model and speculative-grade transition model.

**Comparison between one-year transition probability model and instantaneous transition probability model for complete-grade transitions**

- Comparing to instantaneous transition model, discrete model has less over-dispersion factor. The two-covariate instantaneous transition model even fails to reduce dispersion from one-covariate model.

- From one-covariate results, long interest rates have more significant impact than short rates on instantaneous transition probabilities with stability-migration effect or upgrade-downgrade effect. On the contrary, one-year transition probabilities are more affected by short interest rates than long interest rates by their upgrade-downgrade impact.

**Comparison between one-year transition probability models on investment-grade transitions and speculative-grade transitions**

- Investment-grade transitions are more sensitive to long interest rate risk while speculative-grade transitions are more sensitive to short rate risk.

- For investment-grade transitions, SPREAD is not as important as it is for investment-grade transitions for both stability-migration effect and upgrade-downgrade effect investigation. This maybe because SPREAD is a credit risk measurement for investment-graded BBB rating in Moody’s.
Table 7.1: Estimated one-year transition matrix under the discrete null model

<table>
<thead>
<tr>
<th>Initial rating</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
<th>D</th>
<th>N.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>89.61</td>
<td>6.61</td>
<td>0.41</td>
<td>0.10</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.24</td>
</tr>
<tr>
<td>AA</td>
<td>0.58</td>
<td>88.65</td>
<td>6.55</td>
<td>0.61</td>
<td>0.05</td>
<td>0.11</td>
<td>0.02</td>
<td>0.01</td>
<td>3.42</td>
</tr>
<tr>
<td>A</td>
<td>0.06</td>
<td>2.28</td>
<td>87.48</td>
<td>4.72</td>
<td>0.47</td>
<td>0.21</td>
<td>0.01</td>
<td>0.04</td>
<td>4.73</td>
</tr>
<tr>
<td>BBB</td>
<td>0.03</td>
<td>0.24</td>
<td>5.05</td>
<td>83.04</td>
<td>4.33</td>
<td>0.79</td>
<td>0.12</td>
<td>0.21</td>
<td>6.18</td>
</tr>
<tr>
<td>BB</td>
<td>0.03</td>
<td>0.10</td>
<td>0.43</td>
<td>6.43</td>
<td>74.68</td>
<td>7.13</td>
<td>0.99</td>
<td>0.91</td>
<td>9.30</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.11</td>
<td>0.28</td>
<td>0.49</td>
<td>5.36</td>
<td>73.80</td>
<td>3.48</td>
<td>5.16</td>
<td>11.33</td>
</tr>
<tr>
<td>CCC</td>
<td>0.14</td>
<td>0.00</td>
<td>0.28</td>
<td>1.12</td>
<td>1.54</td>
<td>9.13</td>
<td>53.09</td>
<td>20.93</td>
<td>13.76</td>
</tr>
</tbody>
</table>

\(^1\)Row ratings are initial rating at year beginning; column ratings are final rating at year end
\(^2\)Five zero probabilities: AAA → B, AAA → CCC, AAA → D, B → AAA and CCC → AA
Table 7.2: Discrete one-covariate stability-migration model results for complete-grade transitions

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{e_1}$</th>
<th>$\hat{\beta}(10^{-2})$</th>
<th>$sd(10^{-2})$</th>
<th>$t$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>6.04</td>
<td>5.93</td>
<td>2.41</td>
<td>2.46</td>
<td>2.72</td>
</tr>
<tr>
<td>CPI</td>
<td>0.30</td>
<td>0.36</td>
<td>0.66</td>
<td>0.55</td>
<td>2.76</td>
</tr>
<tr>
<td>PPI</td>
<td>4.06</td>
<td>-1.33</td>
<td>0.66</td>
<td>-2.02</td>
<td>2.70</td>
</tr>
<tr>
<td>PAYRL</td>
<td>3.21</td>
<td>1.65</td>
<td>0.92</td>
<td>1.79</td>
<td>2.74</td>
</tr>
<tr>
<td>UNEMP</td>
<td>4.09</td>
<td>-1.49</td>
<td>0.75</td>
<td>-1.99</td>
<td>2.71</td>
</tr>
<tr>
<td>R3M</td>
<td>6.49</td>
<td>-1.16</td>
<td>0.46</td>
<td>-2.52</td>
<td>2.69</td>
</tr>
<tr>
<td>R6M</td>
<td>5.96</td>
<td>-1.12</td>
<td>0.46</td>
<td>-2.43</td>
<td>2.69</td>
</tr>
<tr>
<td>R1Y</td>
<td>5.73</td>
<td>-1.15</td>
<td>0.48</td>
<td>-2.40</td>
<td>2.69</td>
</tr>
<tr>
<td>GS05</td>
<td>5.55</td>
<td>-1.04</td>
<td>0.45</td>
<td>-2.26</td>
<td>2.69</td>
</tr>
<tr>
<td>GS10</td>
<td>6.34</td>
<td>-1.14</td>
<td>0.46</td>
<td>-2.48</td>
<td>2.69</td>
</tr>
<tr>
<td>GS30</td>
<td>6.09</td>
<td>-1.21</td>
<td>0.50</td>
<td>-2.42</td>
<td>2.69</td>
</tr>
<tr>
<td>SPREAD</td>
<td>16.36</td>
<td>-10.21</td>
<td>2.52</td>
<td>-4.05</td>
<td>2.63</td>
</tr>
</tbody>
</table>

1 business covariate time series, $t =$ year 1, 2, 4, ..., 19
2 improvement in scaled-deviance
3 maximum likelihood estimate of $\hat{\beta}$
4 dispersion inflated standard deviation of $\hat{\beta}$
5 $t$ statistics $t = \hat{\beta}/sd$
6 estimated dispersion factor: $\hat{\phi} = \chi^2/df$ with $df = 1012$
Table 7.3: Discrete one-covariate upgrade-downgrade model results for complete-grade

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{s_t}$</th>
<th>$\beta(10^{-2})$</th>
<th>$sd(10^{-2})$</th>
<th>$t$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>30.70</td>
<td>24.28</td>
<td>4.39</td>
<td>5.53</td>
<td>2.66</td>
</tr>
<tr>
<td>CPI</td>
<td>18.50</td>
<td>-5.08</td>
<td>1.18</td>
<td>-4.31</td>
<td>2.72</td>
</tr>
<tr>
<td>PPI</td>
<td>9.53</td>
<td>-3.79</td>
<td>1.23</td>
<td>-3.08</td>
<td>2.71</td>
</tr>
<tr>
<td>PAYRL</td>
<td>40.60</td>
<td>10.66</td>
<td>1.67</td>
<td>6.38</td>
<td>2.65</td>
</tr>
<tr>
<td>UNEMP</td>
<td>8.66</td>
<td>-4.00</td>
<td>1.36</td>
<td>-2.94</td>
<td>2.72</td>
</tr>
<tr>
<td>R3M</td>
<td>23.93</td>
<td>-4.10</td>
<td>0.84</td>
<td>-4.88</td>
<td>2.65</td>
</tr>
<tr>
<td>R6M</td>
<td>23.51</td>
<td>-4.11</td>
<td>0.85</td>
<td>-4.84</td>
<td>2.66</td>
</tr>
<tr>
<td>R1Y</td>
<td>22.21</td>
<td>-4.19</td>
<td>0.89</td>
<td>-4.71</td>
<td>2.66</td>
</tr>
<tr>
<td>GS05</td>
<td>22.82</td>
<td>-3.89</td>
<td>0.82</td>
<td>-4.74</td>
<td>2.66</td>
</tr>
<tr>
<td>GS10</td>
<td>22.63</td>
<td>-3.98</td>
<td>0.84</td>
<td>-4.74</td>
<td>2.66</td>
</tr>
<tr>
<td>GS30</td>
<td>21.38</td>
<td>-4.18</td>
<td>0.91</td>
<td>-4.59</td>
<td>2.67</td>
</tr>
<tr>
<td>SPREAD</td>
<td>33.26</td>
<td>-27.29</td>
<td>4.74</td>
<td>-5.76</td>
<td>2.64</td>
</tr>
</tbody>
</table>

$^1$d.f. = 1012
Table 7.4: Top ten best discrete two-covariate model results for complete-grade transitions

<table>
<thead>
<tr>
<th></th>
<th>$x_{1,t}$</th>
<th>$x_{2,t}$</th>
<th>$D_{s_1}$</th>
<th>$D_{s_2}$</th>
<th>$\hat{\beta}_1$</th>
<th>$sd_1$</th>
<th>$t_1$</th>
<th>$D_{s_2}$</th>
<th>$\hat{\beta}_2$</th>
<th>$sd_2$</th>
<th>$t_2$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SPREAD</td>
<td>PAYRL</td>
<td>53.87</td>
<td>11.82</td>
<td>-8.66</td>
<td>2.53</td>
<td>-3.42</td>
<td>37.05</td>
<td>9.69</td>
<td>1.61</td>
<td>6.02</td>
<td>2.56</td>
</tr>
<tr>
<td>2</td>
<td>R3M</td>
<td>PAYRL</td>
<td>45.66</td>
<td>4.35</td>
<td>-0.94</td>
<td>0.45</td>
<td>-2.09</td>
<td>38.97</td>
<td>10.22</td>
<td>1.65</td>
<td>6.19</td>
<td>2.61</td>
</tr>
<tr>
<td>3</td>
<td>GS10</td>
<td>PAYRL</td>
<td>45.51</td>
<td>4.16</td>
<td>-0.92</td>
<td>0.46</td>
<td>-2.00</td>
<td>38.98</td>
<td>10.19</td>
<td>1.65</td>
<td>6.18</td>
<td>2.61</td>
</tr>
<tr>
<td>4</td>
<td>GS30</td>
<td>PAYRL</td>
<td>45.30</td>
<td>3.97</td>
<td>-0.97</td>
<td>0.49</td>
<td>-1.98</td>
<td>39.03</td>
<td>10.2</td>
<td>1.65</td>
<td>6.18</td>
<td>2.61</td>
</tr>
<tr>
<td>5</td>
<td>R6M</td>
<td>PAYRL</td>
<td>45.16</td>
<td>3.9</td>
<td>-0.90</td>
<td>0.46</td>
<td>-1.96</td>
<td>39.02</td>
<td>10.23</td>
<td>1.65</td>
<td>6.2</td>
<td>2.61</td>
</tr>
<tr>
<td>6</td>
<td>SPREAD</td>
<td>GDP</td>
<td>45.03</td>
<td>13.2</td>
<td>-9.12</td>
<td>2.51</td>
<td>-3.63</td>
<td>28.23</td>
<td>22.19</td>
<td>4.21</td>
<td>5.27</td>
<td>2.56</td>
</tr>
<tr>
<td>7</td>
<td>R1Y</td>
<td>PAYRL</td>
<td>45.02</td>
<td>3.76</td>
<td>-0.93</td>
<td>0.48</td>
<td>-1.94</td>
<td>39.12</td>
<td>10.24</td>
<td>1.65</td>
<td>6.21</td>
<td>2.61</td>
</tr>
<tr>
<td>8</td>
<td>GS05</td>
<td>PAYRL</td>
<td>44.76</td>
<td>3.5</td>
<td>-0.82</td>
<td>0.44</td>
<td>-1.86</td>
<td>39.04</td>
<td>10.22</td>
<td>1.65</td>
<td>6.19</td>
<td>2.61</td>
</tr>
<tr>
<td>9</td>
<td>PPI</td>
<td>PAYRL</td>
<td>44.29</td>
<td>3.11</td>
<td>-1.15</td>
<td>0.65</td>
<td>-1.77</td>
<td>40.09</td>
<td>10.43</td>
<td>1.65</td>
<td>6.32</td>
<td>2.62</td>
</tr>
<tr>
<td>10</td>
<td>SPREAD</td>
<td>SPREAD</td>
<td>44.12</td>
<td>9.85</td>
<td>-8.04</td>
<td>2.57</td>
<td>-3.13</td>
<td>27.28</td>
<td>-23.78</td>
<td>4.62</td>
<td>-5.15</td>
<td>2.56</td>
</tr>
</tbody>
</table>

1. business-covariate with stability-migration effect  
2. business-covariate with upgrade-downgrade effect  
3. improvement in scaled-deviance from null model  
4. improvement in scaled-deviance from corresponding one-covariate upgrade-downgrade model  
5. maximum likelihood estimates of $\hat{\beta}_1$  
6. dispersion-inflated standard deviation of $\hat{\beta}_1$  
7. asymptotic $t$ test for $H_{10}$  
8. improvement in scaled-deviance from corresponding one-covariate stability-migration model  
9. maximum likelihood estimates of $\hat{\beta}_2$  
10. dispersion-inflated standard deviation of $\hat{\beta}_2$  
11. asymptotic $t$ test for $H_{20}$  
12. estimated dispersion factor $\hat{\phi} = \chi^2/1011$
Table 7.5: Discrete one-covariate stability-migration model results for investment-grade transitions

<table>
<thead>
<tr>
<th>( x_t )</th>
<th>( D_{s_1} )</th>
<th>( \hat{\beta}(10^{-2}) )</th>
<th>( sd(10^{-2}) )</th>
<th>( t )</th>
<th>( \phi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>0.01</td>
<td>0.26</td>
<td>3.14</td>
<td>0.08</td>
<td>2.80</td>
</tr>
<tr>
<td>CPI</td>
<td>0.13</td>
<td>0.30</td>
<td>0.83</td>
<td>0.36</td>
<td>2.81</td>
</tr>
<tr>
<td>PPI</td>
<td>4.03</td>
<td>-1.67</td>
<td>0.82</td>
<td>-2.04</td>
<td>2.75</td>
</tr>
<tr>
<td>PAYRL</td>
<td>0.07</td>
<td>-0.31</td>
<td>1.20</td>
<td>-0.26</td>
<td>2.80</td>
</tr>
<tr>
<td>UNEMP</td>
<td>2.62</td>
<td>-1.53</td>
<td>0.95</td>
<td>-1.61</td>
<td>2.75</td>
</tr>
<tr>
<td>R3M</td>
<td>5.77</td>
<td>-1.37</td>
<td>0.56</td>
<td>-2.45</td>
<td>2.72</td>
</tr>
<tr>
<td>R6M</td>
<td>5.96</td>
<td>-1.40</td>
<td>0.57</td>
<td>-2.46</td>
<td>2.72</td>
</tr>
<tr>
<td>R1Y</td>
<td>6.48</td>
<td>-1.53</td>
<td>0.60</td>
<td>-2.55</td>
<td>2.71</td>
</tr>
<tr>
<td>GS05</td>
<td>6.43</td>
<td>-1.41</td>
<td>0.55</td>
<td>-2.56</td>
<td>2.71</td>
</tr>
<tr>
<td>GS10</td>
<td>6.90</td>
<td>-1.50</td>
<td>0.57</td>
<td>-2.63</td>
<td>2.70</td>
</tr>
<tr>
<td>GS30</td>
<td>6.32</td>
<td>-1.56</td>
<td>0.62</td>
<td>-2.52</td>
<td>2.71</td>
</tr>
<tr>
<td>SPREAD</td>
<td>18.96</td>
<td>-13.49</td>
<td>3.06</td>
<td>-4.41</td>
<td>2.61</td>
</tr>
</tbody>
</table>

1 business covariate time series, \( t = \) year 1, 2, 4 \cdots , 19
2 improvement in scaled-deviance from investment-grade null model
3 maximum likelihood estimate of \( \hat{\beta} \)
4 dispersion inflated standard deviation of \( \hat{\beta} \)
5 \( t \) statistics \( t = \hat{\beta}/sd \)
6 estimated dispersion factor: \( \hat{\phi} = \chi^2/d.f. \) with \( df = 578 \)
Table 7.6: Discrete one-covariate upgrade-downgrade model results for investment-grade transitions

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{s_1}$</th>
<th>$\beta (10^{-2})$</th>
<th>$sd (10^{-2})$</th>
<th>$t$</th>
<th>$\phi^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>23.06</td>
<td>27.76</td>
<td>5.80</td>
<td>4.79</td>
<td>2.71</td>
</tr>
<tr>
<td>CPI</td>
<td>17.75</td>
<td>-6.44</td>
<td>1.53</td>
<td>-4.21</td>
<td>2.77</td>
</tr>
<tr>
<td>PPI</td>
<td>6.05</td>
<td>-3.94</td>
<td>1.60</td>
<td>-2.46</td>
<td>2.79</td>
</tr>
<tr>
<td>PAYRL</td>
<td>25.73</td>
<td>11.17</td>
<td>2.21</td>
<td>5.05</td>
<td>2.73</td>
</tr>
<tr>
<td>UNEMP</td>
<td>15.48</td>
<td>-7.03</td>
<td>1.79</td>
<td>-3.93</td>
<td>2.75</td>
</tr>
<tr>
<td>R3M</td>
<td>16.85</td>
<td>-4.48</td>
<td>1.09</td>
<td>-4.11</td>
<td>2.73</td>
</tr>
<tr>
<td>R6M</td>
<td>17.18</td>
<td>-4.57</td>
<td>1.10</td>
<td>-4.15</td>
<td>2.73</td>
</tr>
<tr>
<td>R1Y</td>
<td>17.07</td>
<td>-4.77</td>
<td>1.16</td>
<td>-4.11</td>
<td>2.72</td>
</tr>
<tr>
<td>GS05</td>
<td>22.60</td>
<td>-5.04</td>
<td>1.06</td>
<td>-4.75</td>
<td>2.70</td>
</tr>
<tr>
<td>GS10</td>
<td>24.10</td>
<td>-5.36</td>
<td>1.10</td>
<td>-4.87</td>
<td>2.70</td>
</tr>
<tr>
<td>GS30</td>
<td>24.60</td>
<td>-5.85</td>
<td>1.19</td>
<td>-4.92</td>
<td>2.70</td>
</tr>
<tr>
<td>SPREAD</td>
<td>22.37</td>
<td>-29.08</td>
<td>6.16</td>
<td>-4.72</td>
<td>2.70</td>
</tr>
</tbody>
</table>

$^1$ d.f. = 578
Table 7.7: Top ten best two-covariate model results for investment-grade transitions

<table>
<thead>
<tr>
<th>$x_{1,t}$</th>
<th>$x_{2,t}$</th>
<th>$D_{s3}$</th>
<th>$D_{s21}$</th>
<th>$\beta_1$</th>
<th>$t_1$</th>
<th>$D_{s22}$</th>
<th>$\beta_2$</th>
<th>$sd_{s1}$</th>
<th>$sd_{s2}$</th>
<th>$t_2$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SPREAD</td>
<td>GS30</td>
<td>46.55</td>
<td>20.12</td>
<td>-13.70</td>
<td>3.01</td>
<td>-4.55</td>
<td>26.87</td>
<td>-5.59</td>
<td>1.08</td>
<td>-5.18</td>
</tr>
<tr>
<td>2</td>
<td>SPREAD</td>
<td>GS10</td>
<td>45.96</td>
<td>20.07</td>
<td>-13.69</td>
<td>3.01</td>
<td>-4.55</td>
<td>26.28</td>
<td>-5.10</td>
<td>1.00</td>
<td>-5.10</td>
</tr>
<tr>
<td>3</td>
<td>SPREAD</td>
<td>PAYRL</td>
<td>45.22</td>
<td>17.36</td>
<td>-12.94</td>
<td>3.05</td>
<td>-4.24</td>
<td>25.85</td>
<td>10.15</td>
<td>2.01</td>
<td>5.05</td>
</tr>
<tr>
<td>4</td>
<td>SPREAD</td>
<td>GS05</td>
<td>44.10</td>
<td>19.86</td>
<td>-13.64</td>
<td>3.02</td>
<td>-4.52</td>
<td>24.50</td>
<td>-4.77</td>
<td>0.97</td>
<td>-4.92</td>
</tr>
<tr>
<td>5</td>
<td>SPREAD</td>
<td>SPREAD</td>
<td>43.65</td>
<td>19.64</td>
<td>-13.56</td>
<td>3.02</td>
<td>-4.49</td>
<td>24.04</td>
<td>-27.10</td>
<td>5.56</td>
<td>-4.87</td>
</tr>
<tr>
<td>6</td>
<td>SPREAD</td>
<td>GDP</td>
<td>42.88</td>
<td>18.20</td>
<td>-13.08</td>
<td>3.03</td>
<td>-4.32</td>
<td>23.36</td>
<td>25.52</td>
<td>5.32</td>
<td>4.80</td>
</tr>
<tr>
<td>7</td>
<td>SPREAD</td>
<td>R6M</td>
<td>38.15</td>
<td>19.71</td>
<td>-13.63</td>
<td>3.03</td>
<td>-4.50</td>
<td>18.70</td>
<td>-4.32</td>
<td>1.00</td>
<td>-4.32</td>
</tr>
<tr>
<td>8</td>
<td>SPREAD</td>
<td>R1Y</td>
<td>38.09</td>
<td>19.79</td>
<td>-13.65</td>
<td>3.03</td>
<td>-4.50</td>
<td>18.64</td>
<td>-4.53</td>
<td>1.05</td>
<td>-4.31</td>
</tr>
<tr>
<td>9</td>
<td>SPREAD</td>
<td>R3M</td>
<td>37.84</td>
<td>19.74</td>
<td>-13.63</td>
<td>3.03</td>
<td>-4.50</td>
<td>18.38</td>
<td>-4.25</td>
<td>0.99</td>
<td>-4.29</td>
</tr>
<tr>
<td>10</td>
<td>SPREAD</td>
<td>CPI</td>
<td>36.98</td>
<td>17.94</td>
<td>-13.14</td>
<td>3.06</td>
<td>-4.29</td>
<td>17.83</td>
<td>-5.69</td>
<td>1.36</td>
<td>-4.18</td>
</tr>
</tbody>
</table>

1. business-covariate with stability-migration effect
2. business-covariate with upgrade-downgrade effect
3. improvement in scaled-deviance from investment-grade null model
4. improvement in scaled-deviance from corresponding investment-grade one-covariate upgrade-downgrade model
5. investment-grade maximum likelihood estimates of $\beta_1$
6. dispersion-inflated standard deviation of $\hat{\beta}_1$
7. asymptotic $t$ test for $H_{10}$
8. improvement in scaled-deviance from corresponding investment-grade one-covariate stability-migration model
9. investment-grade maximum likelihood estimates of $\beta_2$
10. dispersion-inflated standard deviation of $\hat{\beta}_2$
11. asymptotic $t$ test for $H_{20}$
12. estimated dispersion factor $\hat{\phi} = \chi^2 / d.f., d.f. = 577$
Table 7.8: Discrete one-covariate stability-migration model results for speculative-grade transitions

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{s_1}$</th>
<th>$\beta(10^{-2})$</th>
<th>$sd(10^{-2})$</th>
<th>$t$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>15.86</td>
<td>14.93</td>
<td>3.71</td>
<td>4.02</td>
<td>2.55</td>
</tr>
<tr>
<td>CPI</td>
<td>0.18</td>
<td>0.46</td>
<td>1.04</td>
<td>0.44</td>
<td>2.70</td>
</tr>
<tr>
<td>PPI</td>
<td>0.44</td>
<td>-0.73</td>
<td>1.05</td>
<td>-0.70</td>
<td>2.66</td>
</tr>
<tr>
<td>PAYRL</td>
<td>11.22</td>
<td>4.86</td>
<td>1.43</td>
<td>3.40</td>
<td>2.58</td>
</tr>
<tr>
<td>UNEMP</td>
<td>1.46</td>
<td>-1.42</td>
<td>1.08</td>
<td>-1.31</td>
<td>2.66</td>
</tr>
<tr>
<td>R3M</td>
<td>1.05</td>
<td>-0.78</td>
<td>0.72</td>
<td>-1.08</td>
<td>2.65</td>
</tr>
<tr>
<td>R6M</td>
<td>0.63</td>
<td>-0.61</td>
<td>0.73</td>
<td>-0.84</td>
<td>2.66</td>
</tr>
<tr>
<td>R1Y</td>
<td>0.33</td>
<td>-0.46</td>
<td>0.76</td>
<td>-0.61</td>
<td>2.67</td>
</tr>
<tr>
<td>GS05</td>
<td>0.30</td>
<td>-0.40</td>
<td>0.68</td>
<td>-0.59</td>
<td>2.68</td>
</tr>
<tr>
<td>GS10</td>
<td>0.49</td>
<td>-0.52</td>
<td>0.69</td>
<td>-0.75</td>
<td>2.67</td>
</tr>
<tr>
<td>GS30</td>
<td>0.59</td>
<td>-0.62</td>
<td>0.74</td>
<td>-0.84</td>
<td>2.67</td>
</tr>
<tr>
<td>SPREAD</td>
<td>0.87</td>
<td>-4.02</td>
<td>4.01</td>
<td>-1.00</td>
<td>2.66</td>
</tr>
</tbody>
</table>

$^1$d.f. = 433
Table 7.9: Discrete one-covariate upgrade-downgrade model results for speculative-grade transitions

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{s_t}$</th>
<th>$\beta(10^{-2})$</th>
<th>$sd(10^{-2})$</th>
<th>$t$</th>
<th>$\phi^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>8.26</td>
<td>19.38</td>
<td>7.00</td>
<td>2.77</td>
<td>2.61</td>
</tr>
<tr>
<td>CPI</td>
<td>2.54</td>
<td>-2.98</td>
<td>2.06</td>
<td>-1.45</td>
<td>2.68</td>
</tr>
<tr>
<td>PPI</td>
<td>3.45</td>
<td>-3.56</td>
<td>2.26</td>
<td>-1.58</td>
<td>2.62</td>
</tr>
<tr>
<td>PAYRL</td>
<td>14.75</td>
<td>9.90</td>
<td>2.65</td>
<td>3.74</td>
<td>2.56</td>
</tr>
<tr>
<td>UNEMP</td>
<td>0.02</td>
<td>0.27</td>
<td>2.83</td>
<td>0.10</td>
<td>2.70</td>
</tr>
<tr>
<td>R3M</td>
<td>7.14</td>
<td>-3.53</td>
<td>1.58</td>
<td>-2.23</td>
<td>2.57</td>
</tr>
<tr>
<td>R6M</td>
<td>6.50</td>
<td>-3.41</td>
<td>1.61</td>
<td>-2.12</td>
<td>2.58</td>
</tr>
<tr>
<td>R1Y</td>
<td>5.53</td>
<td>-3.29</td>
<td>1.70</td>
<td>-1.94</td>
<td>2.59</td>
</tr>
<tr>
<td>GS05</td>
<td>2.86</td>
<td>-2.17</td>
<td>1.70</td>
<td>-1.28</td>
<td>2.63</td>
</tr>
<tr>
<td>GS10</td>
<td>2.22</td>
<td>-1.95</td>
<td>1.78</td>
<td>-1.10</td>
<td>2.64</td>
</tr>
<tr>
<td>GS30</td>
<td>1.52</td>
<td>-1.74</td>
<td>1.94</td>
<td>-0.90</td>
<td>2.65</td>
</tr>
<tr>
<td>SPREAD</td>
<td>10.85</td>
<td>-24.52</td>
<td>9.91</td>
<td>-2.47</td>
<td>2.56</td>
</tr>
</tbody>
</table>

$^1d.f. = 433$
Table 7.10: Top ten best discrete two-covariate model results for speculative-grade transitions

<table>
<thead>
<tr>
<th>$x_{1,t}$</th>
<th>$x_{2,t}$</th>
<th>$D_{s3}$</th>
<th>$D_{s21}$</th>
<th>$\hat{\beta}_1$</th>
<th>sd$_1$</th>
<th>$t_1$</th>
<th>$D_{s22}$</th>
<th>$\hat{\beta}_2$</th>
<th>sd$_2$</th>
<th>$t_2$</th>
<th>$\phi^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>SPREAD</td>
<td>23.53</td>
<td>12.27</td>
<td>13.11</td>
<td>3.80</td>
<td>3.45</td>
<td>7.17</td>
<td>-19.27</td>
<td>9.11</td>
<td>-2.12</td>
<td>2.47</td>
</tr>
<tr>
<td>GDP</td>
<td>PAYRL</td>
<td>22.47</td>
<td>7.35</td>
<td>11.06</td>
<td>4.06</td>
<td>2.72</td>
<td>6.30</td>
<td>6.70</td>
<td>2.74</td>
<td>2.45</td>
<td>2.50</td>
</tr>
<tr>
<td>GDP</td>
<td>R3M</td>
<td>20.56</td>
<td>13.15</td>
<td>13.59</td>
<td>3.76</td>
<td>3.61</td>
<td>4.27</td>
<td>-2.65</td>
<td>1.48</td>
<td>-1.79</td>
<td>2.48</td>
</tr>
<tr>
<td>GDP</td>
<td>R6M</td>
<td>20.06</td>
<td>13.31</td>
<td>13.68</td>
<td>3.76</td>
<td>3.64</td>
<td>3.81</td>
<td>-2.53</td>
<td>1.51</td>
<td>-1.68</td>
<td>2.49</td>
</tr>
<tr>
<td>GDP</td>
<td>R1Y</td>
<td>19.38</td>
<td>13.63</td>
<td>13.84</td>
<td>3.76</td>
<td>3.68</td>
<td>3.18</td>
<td>-2.42</td>
<td>1.59</td>
<td>-1.52</td>
<td>2.49</td>
</tr>
<tr>
<td>PAYRL</td>
<td>SPREAD</td>
<td>18.91</td>
<td>7.78</td>
<td>4.06</td>
<td>1.49</td>
<td>2.72</td>
<td>7.33</td>
<td>-19.76</td>
<td>9.36</td>
<td>-2.11</td>
<td>2.50</td>
</tr>
<tr>
<td>GDP</td>
<td>PPI</td>
<td>18.83</td>
<td>15.22</td>
<td>14.48</td>
<td>3.69</td>
<td>3.92</td>
<td>2.67</td>
<td>-3.03</td>
<td>2.15</td>
<td>-1.41</td>
<td>2.50</td>
</tr>
<tr>
<td>PAYRL</td>
<td>PAYRL</td>
<td>18.13</td>
<td>3.18</td>
<td>2.95</td>
<td>1.66</td>
<td>1.78</td>
<td>6.68</td>
<td>7.26</td>
<td>2.94</td>
<td>2.47</td>
<td>2.53</td>
</tr>
<tr>
<td>GDP</td>
<td>GDP</td>
<td>17.60</td>
<td>9.10</td>
<td>12.63</td>
<td>4.16</td>
<td>3.04</td>
<td>1.66</td>
<td>9.25</td>
<td>7.33</td>
<td>1.26</td>
<td>2.53</td>
</tr>
<tr>
<td>GDP</td>
<td>GS05</td>
<td>17.33</td>
<td>14.34</td>
<td>14.26</td>
<td>3.78</td>
<td>3.77</td>
<td>1.31</td>
<td>-1.43</td>
<td>1.57</td>
<td>-0.91</td>
<td>2.52</td>
</tr>
</tbody>
</table>

$^1$ d.f. = 432
Table 7.11: Estimated instantaneous transition matrix under the null model

<table>
<thead>
<tr>
<th>Initial rating</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
<th>D</th>
<th>N.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>-0.97</td>
<td>0.64</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.30</td>
</tr>
<tr>
<td>AA</td>
<td>0.06</td>
<td>-1.12</td>
<td>0.68</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.32</td>
</tr>
<tr>
<td>A</td>
<td>0.01</td>
<td>0.23</td>
<td>-1.27</td>
<td>0.54</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.43</td>
</tr>
<tr>
<td>BBB</td>
<td>0</td>
<td>0.03</td>
<td>0.53</td>
<td>-1.87</td>
<td>0.63</td>
<td>0.09</td>
<td>0</td>
<td>0.03</td>
<td>0.56</td>
</tr>
<tr>
<td>BB</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
<td>0.74</td>
<td>-2.90</td>
<td>1.06</td>
<td>0.09</td>
<td>0.07</td>
<td>0.87</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.06</td>
<td>0.64</td>
<td>-2.91</td>
<td>0.81</td>
<td>0.35</td>
<td>1.05</td>
</tr>
<tr>
<td>CCC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.34</td>
<td>1.11</td>
<td>-5.89</td>
<td>2.77</td>
<td>1.67</td>
</tr>
<tr>
<td>D</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>N.R.</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>
</tbody>
</table>

1Number 0 is real zero in instantaneous model
Table 7.12: Instantaneous-compounded one-year transition matrix approximation under the continuous null model

<table>
<thead>
<tr>
<th>Initial rating</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
<th>D</th>
<th>N.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>89.04</td>
<td>6.82</td>
<td>0.55</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.55</td>
</tr>
<tr>
<td>AA</td>
<td>0.68</td>
<td>87.49</td>
<td>7.08</td>
<td>0.84</td>
<td>0.05</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>3.85</td>
</tr>
<tr>
<td>A</td>
<td>0.09</td>
<td>2.40</td>
<td>86.15</td>
<td>5.47</td>
<td>0.74</td>
<td>0.07</td>
<td>0.00</td>
<td>0.01</td>
<td>5.06</td>
</tr>
<tr>
<td>BBB</td>
<td>0.00</td>
<td>0.39</td>
<td>5.29</td>
<td>80.48</td>
<td>5.71</td>
<td>1.13</td>
<td>0.07</td>
<td>0.34</td>
<td>6.58</td>
</tr>
<tr>
<td>BB</td>
<td>0.00</td>
<td>0.02</td>
<td>0.84</td>
<td>6.74</td>
<td>71.26</td>
<td>9.08</td>
<td>1.01</td>
<td>1.14</td>
<td>9.90</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83</td>
<td>5.58</td>
<td>71.38</td>
<td>5.78</td>
<td>4.70</td>
<td>11.68</td>
</tr>
<tr>
<td>CCC</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.16</td>
<td>2.79</td>
<td>8.09</td>
<td>49.70</td>
<td>24.09</td>
<td>15.16</td>
</tr>
</tbody>
</table>
Table 7.13: Continuous one-covariate stability-migration model results

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{x_1}$</th>
<th>$\beta$</th>
<th>$\phi^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>2.08</td>
<td>0.10</td>
<td>9.59</td>
</tr>
<tr>
<td>CPI</td>
<td>0.06</td>
<td>0.00</td>
<td>9.79</td>
</tr>
<tr>
<td>PPI</td>
<td>1.45</td>
<td>-0.02</td>
<td>9.77</td>
</tr>
<tr>
<td>PAYRL</td>
<td>1.16</td>
<td>0.03</td>
<td>9.71</td>
</tr>
<tr>
<td>UNEMP</td>
<td>3.12</td>
<td>-0.04</td>
<td>10.29</td>
</tr>
<tr>
<td>R3M</td>
<td>3.23</td>
<td>-0.03</td>
<td>9.89</td>
</tr>
<tr>
<td>R6M</td>
<td>3.13</td>
<td>-0.03</td>
<td>9.89</td>
</tr>
<tr>
<td>R1Y</td>
<td>3.15</td>
<td>-0.03</td>
<td>9.91</td>
</tr>
<tr>
<td>GS05</td>
<td>3.52</td>
<td>-0.03</td>
<td>10.02</td>
</tr>
<tr>
<td>GS10</td>
<td>3.97</td>
<td>-0.03</td>
<td>10.06</td>
</tr>
<tr>
<td>GS30</td>
<td>3.95</td>
<td>-0.03</td>
<td>10.09</td>
</tr>
<tr>
<td>SPREAD</td>
<td>7.07</td>
<td>-0.21</td>
<td>10.17</td>
</tr>
</tbody>
</table>

$d.f. = 1030$
Table 7.14: Continuous one-covariate upgrade-downgrade model results

<table>
<thead>
<tr>
<th>$x_t$</th>
<th>$D_{x_t}$</th>
<th>$\beta$</th>
<th>$\phi^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>7.53</td>
<td>0.19</td>
<td>9.82</td>
</tr>
<tr>
<td>CPI</td>
<td>4.40</td>
<td>-0.04</td>
<td>9.76</td>
</tr>
<tr>
<td>PPI</td>
<td>2.00</td>
<td>-0.02</td>
<td>9.70</td>
</tr>
<tr>
<td>PAYRL</td>
<td>9.25</td>
<td>0.09</td>
<td>10.18</td>
</tr>
<tr>
<td>UNEMP</td>
<td>3.37</td>
<td>-0.04</td>
<td>9.49</td>
</tr>
<tr>
<td>R3M</td>
<td>6.47</td>
<td>-0.04</td>
<td>9.71</td>
</tr>
<tr>
<td>R6M</td>
<td>6.43</td>
<td>-0.04</td>
<td>9.67</td>
</tr>
<tr>
<td>R1Y</td>
<td>6.23</td>
<td>-0.04</td>
<td>9.61</td>
</tr>
<tr>
<td>GS05</td>
<td>6.75</td>
<td>-0.03</td>
<td>9.44</td>
</tr>
<tr>
<td>GS10</td>
<td>6.88</td>
<td>-0.04</td>
<td>9.39</td>
</tr>
<tr>
<td>GS30</td>
<td>6.61</td>
<td>-0.04</td>
<td>9.38</td>
</tr>
<tr>
<td>SPREAD</td>
<td>9.68</td>
<td>-0.24</td>
<td>9.52</td>
</tr>
</tbody>
</table>

$^1$d.f. = 1030
Table 7.15: Top ten best continuous two-covariate model results

<table>
<thead>
<tr>
<th>$x_{1,t}$</th>
<th>$x_{2,t}$</th>
<th>$D_{s3}$</th>
<th>$D_{s21}$</th>
<th>$D_{s22}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\phi^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SPREAD</td>
<td>26.45</td>
<td>18.89</td>
<td>20.68</td>
<td>-0.13</td>
<td>0.09</td>
<td>12.45</td>
</tr>
<tr>
<td>2</td>
<td>SPREAD</td>
<td>26.27</td>
<td>18.27</td>
<td>20.03</td>
<td>-0.12</td>
<td>-0.22</td>
<td>11.51</td>
</tr>
<tr>
<td>3</td>
<td>GDP</td>
<td>26.07</td>
<td>17.92</td>
<td>24.31</td>
<td>0.08</td>
<td>-0.24</td>
<td>11.31</td>
</tr>
<tr>
<td>4</td>
<td>SPREAD</td>
<td>25.96</td>
<td>19.75</td>
<td>19.93</td>
<td>-0.14</td>
<td>0.20</td>
<td>11.92</td>
</tr>
<tr>
<td>5</td>
<td>GDP</td>
<td>25.34</td>
<td>17.74</td>
<td>23.73</td>
<td>0.06</td>
<td>0.10</td>
<td>12.40</td>
</tr>
<tr>
<td>6</td>
<td>PAYRL</td>
<td>25.16</td>
<td>17.13</td>
<td>24.18</td>
<td>0.02</td>
<td>-0.25</td>
<td>11.48</td>
</tr>
<tr>
<td>7</td>
<td>R3M</td>
<td>25.14</td>
<td>17.12</td>
<td>22.36</td>
<td>-0.01</td>
<td>-0.24</td>
<td>11.48</td>
</tr>
<tr>
<td>8</td>
<td>CPI</td>
<td>25.14</td>
<td>17.27</td>
<td>25.09</td>
<td>0.02</td>
<td>-0.28</td>
<td>11.71</td>
</tr>
<tr>
<td>9</td>
<td>R6M</td>
<td>25.07</td>
<td>17.05</td>
<td>22.38</td>
<td>-0.01</td>
<td>-0.24</td>
<td>11.49</td>
</tr>
<tr>
<td>10</td>
<td>GS10</td>
<td>25.07</td>
<td>17.04</td>
<td>21.60</td>
<td>-0.01</td>
<td>-0.24</td>
<td>11.51</td>
</tr>
</tbody>
</table>

$^1$ d.f. = 1029
Chapter 8

Summary and Future Work

8.1 Summary

In this dissertation, we have explored the macro-economic credit cycle and studied debt-issuers’ credit risk drift process under the impact of business conditions.

In the first part of this research, we applied Complex Singular Value Decomposition (CSVD) analysis on all credit condition indicators and business condition indicators for the U.S. economy during 1987 – 1999. From the first order CSVD, we proved the existence of a credit cycle and the leading-lagging relation between business cycle and credit cycle. The conclusion in our Part I study provided the foundation, namely that credit conditions can be predicted by business conditions.

Motivated by the Part I result, to model an issuer’s credit risk indicated by its credit rating, we investigated the extrinsic business condition impact on the intrinsic rating transition probabilities. The business condition was known and represented by the business variables in Part I study. Two business impacts were examined: upgrade-
downgrade effect and stability-migration effect. Under an independent Markov rating transition process assumption, we proposed a Generalized Linear Model (GLM) of rating transition probabilities representing intrinsic transition probabilities as baselines which were adjusted by one or two business impacts. We proposed two similar GLMs: one is a one-year discrete transition probability model, the other is an instantaneous transition probability model. The latter is more practical in deriving transition probabilities over an arbitrary/continuous time interval.

Our Part II result showed the significance of a business impact on issuers’ credit rating transitions in both discrete and continuous time frameworks. Specifically, under good business condition, all ratings tend to be more stable and rating transitions tend to give more upgrades; under bad business conditions, all ratings tend to be less stable and rating transitions tend to show more downgrades. The result also showed investment-grade transitions are more sensitive to long interest rate risk and speculative-grade transitions are more sensitive to short interest rate risk. The continuous model has worse fit than the discrete model in terms of its greater over-dispersion factor.

## 8.2 Future work

There are two major discussions we left for future work. One is to model the credit cycle using business variables. All the credit variables used in CSVD analysis can make up an index to describe the credit conditions. In an economic cycle, represented by credit conditions and business conditions, all credit variables clustering together suggests that we can model this credit index by other business time series.
One possible model candidate would be a general transform function which allows the credit index to depend on current and past values of the explanatory business variables.

The other improvement on this study would be the treatment of business conditions. In this dissertation, the business covariate is assumed to be known. In fact, all the transition probabilities are conditional probabilities on this known factor. For the purpose of predicting credit rating transition probabilities, usually future business conditions are unknown. With the knowledge of a stochastic process for a business covariate, the marginal transition distribution can be obtained by integrating the conditional transition probability over the distribution space of this business covariate. However, this promising approach will involve a lot of computational considerations.
Bibliography


