Credit Risk Modeling and Analysis Using Copula Method and Changepoint Approach to Survival Data

Bo Qian

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Credit Risk Modeling and Analysis Using Copula Method and Changepoint Approach to Survival Data

Bo Qian

This thesis consists of two parts. The first part uses Gaussian Copula and Student's t Copula as the main tools to model the credit risk in securitizations and re-securitizations. The second part proposes a statistical procedure to identify changepoints in Cox model of survival data.

The recent 2007-2009 financial crisis has been regarded as the worst financial crisis since the Great Depression by leading economists. The securitization sector took a lot of blame for the crisis because of the connection of the securitized products created from mortgages to the collapse of the housing market. The first part of this thesis explores the relationship between securitized mortgage products and the 2007-2009 financial crisis using the Copula method as the main tool. We show in this part how loss distributions of securitizations and re-securitizations can be derived or calculated in a new model. Simulations are conducted to examine the

effectiveness of the model. As an application, the model is also used to examine whether and where the ratings of securitized products could be flawed.

On the other hand, the lag effect and saturation effect problems are common and important problems in survival data analysis. They belong to a general class of problems where the treatment effect takes occasional jumps instead of staying constant throughout time. Therefore, they are essentially the changepoint problems in statistics. The second part of this thesis focuses on extending Lai and Xing's recent work in changepoint modeling, which was developed under a time series and Bayesian setup, to the lag effect problems in survival data. A general changepoint approach for Cox model is developed. Simulations and real data analyses are conducted to illustrate the effectiveness of the procedure and how it should be implemented and interpreted.

Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my deepest gratitude toward my Ph.D advisor, Professor Zhiliang Ying, from whom I have received so much advice, constant support, and constructive criticism on research. I would like to thank Professor Haipeng Xing for his kind help on the changepoint model, and thank Professor Tian Zheng, Victor de la Pena, Quanshui Zhao, and Yixin Fang for their insightful comments.

In addition, I am deeply thankful to my parents, whose love has supported me during my life in the States. I would like to give special thanks to my wife, Yi-Hsuan Lee, without whose patience and encouragement I would not have been able to complete this thesis.

Bo Qian

May 2013

Chapter 1

Introduction

This thesis consists of two parts. The first part utilizes Gaussian Copula and Student's t Copula to model the credit risk in securitizations and re-securitizations. The second part proposes a statistical procedure to identify changepoints in Cox model of survival data.

1.1 Credit Risk Modeling and Analysis Using Copula Method

The financial crisis of 2007-2009 has been regarded as the worst financial crisis since the Great Depression by leading economists. During a two-year span, we have witnessed the whole US financial system and all the institutions going through extraordinary changes. In 2009 alone, 140 banks failed, compared to a total of 3 failures from 2005 to 2007. The crisis also saw millions of home owners losing their houses to foreclosures. The majority of these mortgages are so-called subprime or Alt-A, since the borrowers tend to be of lesser credit worthiness as opposed to prime borrowers. Therefore, the 2007-2009 crisis is often referred to as the 'subprime crisis' as well.

The securitization sector in the financial industry took a lot of blame during and after the subprime crisis and credit crunch. This sector is very closely connected to mortgages because of the enormous number of securitizations and re-securitizations created from residential mortgages before and during the crisis. These securities, often referred to as residential mortgage-backed securities (RMBS) and collateralized debt obligations (CDO) backed by RMBS tranches, were in the center of this turmoil and were direct or indirect causes of many of the bank failures. They relate to plenty of factors that are generally deemed to have contributed to the crisis, e.g., subprime lending, housing bubble, complexity and lack of transparency of financial products, incompetence of credit ratings, misuse or misunderstanding of profound mathematical models, etc.

The first part of this thesis is largely an effort to explore the relationship between these securitized products and the subprime crisis using Copula method as the main tool, attempting to answer the questions of what kind of role they played in the crisis. We show in this part how loss distributions of securitizations and re-securitizations can be derived or calculated in a new model. As an application, the model is also used to examine whether and where the ratings of securitized products could be flawed.

1.2 Changepoint Approach to Survival Data

Cox model is the most widely used regression model in survival data analysis. It models failure time by decomposing the hazard rate function into two parts: the baseline hazard function, describing how hazard (risk) changes over time at 'baseline' levels of covariates; and the effect parameters, describing how the hazard varies in response to explanatory covariates.

In survival analysis, it is often not unreasonable to assume that a treatment does not start to affect the risk of failure until after a certain time lag, which can be called "lag effect". There is also the possibility of a "saturation effect", where the treatment stops to affect the risk of failure after a certain period of time. In the terms of Cox model, these situations are equivalent to a time-varying covariate effect parameter, staying at a constant value for a period of time and then taking a jump to a new value. The statistical problem of identifying the location of jumps falls into a general class of problems called changepoint problems.

Changepoint problems have been studied in the statistical literature since the 1970s, and most recently by Lai and Xing (2011). The second part of this thesis extends Lai and Xing's changepoint method to the lag effect problem of survival data and develops a changepoint procedure for Cox model. Simulations and real data analyses are conducted to illustrate the effectiveness of the procedure and how it should be implemented and interpreted.

Chapter 2

Credit Risk Modeling and Analysis Using Copula Method

2.1 Introduction

The financial crisis of 2007-2009 has been regarded as the worst financial crisis since the Great Depression by some leading economists. During a two-year span, we have witnessed the whole US financial system and all the institutions going through extraordinary changes.

In March 2008, the concerns that investment bank Bear Stearns would collapse resulted in its fire-sale to JP Morgan Chase. The crisis hit its peak in September and October 2008. Several major institutions either failed, were acquired under duress, or were subject to government takeover. These included Lehman Brothers, Merrill Lynch, Fannie Mae, Freddie Mac, and AIG. The receivership of Washington Mutual Bank by federal regulators on September 26, 2008, was the largest bank failure in U.S. history. Twenty-five banks in the United States failed and were taken over by the Federal Deposit Insurance Corporation (FDIC) in 2008, after only three failures in 2007 and none in 2006 or 2005. According to CNN, 140 more banks failed in 2009, causing the FDIC's deposit insurance fund to slip into the red for the first time since 1991. In a dramatic meeting on September 18, 2008, Treasury Secretary Henry Paulson and Fed Chairman Ben Bernanke met with key legislators to propose a $700 billion emergency bailout.

The US economic and financial system underwent a huge contraction period during the crisis. According to the Bureau of Economic Analysis, Real Gross Domestic Product, the output of goods and services produced by labor and property located in the United States, decreased at an annual rate of approximately 6 percent in the fourth quarter of 2008 and first quarter of 2009, versus activity in the year-ago periods. According to the Bureau of Labor Statistics, the U.S. unemployment rate increased to 9.8% by September 2009, the highest rate since 1983 and roughly twice the pre-crisis rate. The average hours per work week declined to 33, the lowest level since the government began collecting the data in 1964.

Roger Altman (2009), who was U.S. Deputy Treasury Secretary in 1993-94, wrote that "the crash of 2008 has inflicted profound damage on the US financial system, its economy, and its standing in the world; the crisis is an important geopolitical setback...the crisis has coincided with historical forces that were already shifting the world's focus away from the United States. Over the medium term, the United States will have to operate from a smaller global platform—while others, especially China, will have a chance to rise faster."

Since the start of the crisis, economists and financial researchers have been reflecting and contemplating, trying to figure out the real cause of the crisis behind the scenes. Most people tend to agree that the dramatic rise in mortgage delinquencies and foreclosures was the fuse that set off most of the events afterwards. That is why this past crisis is also referred to as the 'subprime crisis', since most of the delinquencies came from subprime and Alt-A mortgages, which are typically of lesser credit worthiness as opposed to prime mortgages.

The securitization sector in the financial industry took a lot of blame during the subprime crisis and credit crunch. This sector is very closely connected to mortgages because of the enormous number of securitizations and re-securitizations created from residential mortgages before and during the crisis. These securities, often referred to as residential mortgage-backed securities (RMBS) and collateralized debt obligations (CDO) backed by RMBS tranches, were in the center of this turmoil and were direct or indirect causes of many of the bank failures. They relate to plenty of factors that generally deemed to have contributed to the crisis, e.g., subprime lending, housing bubble, complexity and lack of transparency of financial products, incompetence of credit ratings, misuse or misunderstanding of profound mathematical models, etc.

This part of the thesis is largely an effort to explore the relationship between these securitized products and the subprime crisis through a mathematical way, attempting to answer the questions of what kind of role they played in this crisis.

2.1.1 Securitization and Re-securitization

The financial product of main interest in this thesis will be the so-called 'structured products', namely 'securitizations' and 're-securitizations'. They got this name because they are usually structured and 'tranched' in such a way that will give investors different choices for risk and return profile.

As illustrated by Figure 2.1, the typical practice of structuring a re-securitization deal is as follows: firstly, thousands of mortgages (often times of subprime quality) are gathered into one portfolio and securitized. The deal is sliced into multiple tranches which, with different risk and return profile, would cater to the needs of different investors. The top tranche (usually called *super-senior* or *senior* tranche), which is of the highest quality and lowest risk, will often be bought by big institutions like insurance companies. The middle tranches (usually called *mezzanine* tranches) are more risky but generate bigger returns. The lowest tranche is usually called 'equity' tranche for the similarity of its position in a liability structure to a stock. Equity tranches are very risky and are usually sold to hedge funds that seek high returns. This first process is called 'securitization'. The end product is called

'RMBS' in this case.

The second step, called 're-securitization', involves securitizing tens or hundreds of these RMBS tranches (usually mezzanine tranches or subordinate senior tranches) again. This portfolio of RMBS tranches is also sliced into senior, mezzanine and equity tranches and sold to various investors. This kind of product is called a 'CDO of RMBS'.

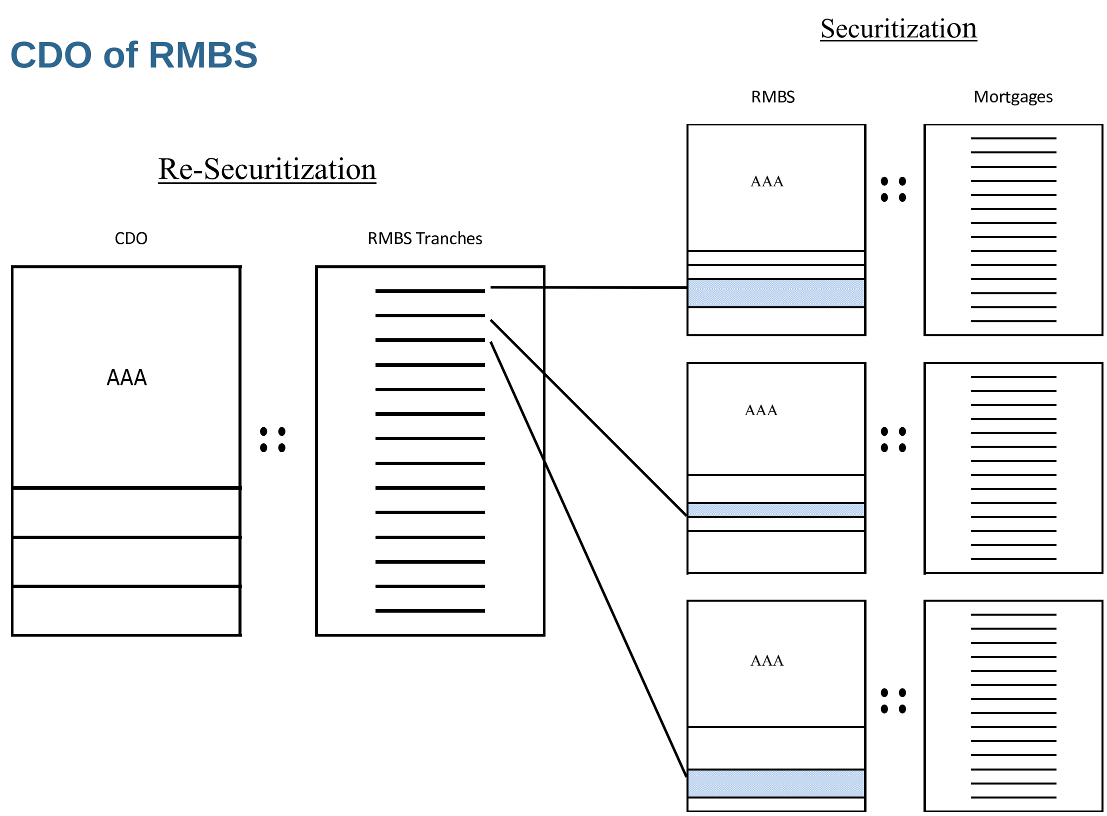
There is about $14.6 trillion in total U.S. mortgage debt outstanding. There are about $8.9 trillion in total U.S. mortgage-related securities. Mortgage-backed securities can be considered to have been in the tens of trillions, if 'Credit Default Swaps' (CDSs) are taken into account. ('Credit Default Swaps' are synthetic instruments that reference a tranche of a securitization or re-securitization.)

In general, securitizations and re-securitizations do not have to be structured from mortgages only. As a matter of fact, people construct them from student loans, car loans, corporate loans, etc. in practice. These products may differ in certain deal specifics from mortgage securities, but the overall ideas of tranching and risk/return tradeoff remain exactly the same.

2.1.2 Economic Capital and Loss Distribution

In finance, mainly for banks and other financial services firms, economic capital is the amount of risk capital, assessed on a realistic basis, which a firm requires to cover the risks of the businesses that it is running as a going concern, such as

Figure 2.1: Securitization and re-securitization



market risk, credit risk, and operational risk. It is the amount of money which is needed to secure survival in a worst case scenario. Firms and financial services regulators should then aim to hold risk capital of an amount equal to or above economic capital.

Dev (2004) introduced the concept of economic capital as follows:

"Economic capital is a measure of risk. It is a single measure that captures the unexpected losses and reduction in value or income from a portfolio or business in a financial institution. The risk arises from the *unexpected* nature of the losses as distinct from *expected* losses, which are considered part of doing business and are covered by reserves and income. Economic capital covers all unexpected events except the catastrophic ones, for which it is impossible to hold capital. Economic capital is a common currency in which all risks of a financial institution can be measured, enabling comparison of risk across different risks, across diverse businesses and across different financial institutions."

Formally, economic capital is often calculated by determining the amount of money that the firm needs to ensure that its balance sheet stays solvent over a certain time period (i.e., the time period is usually one year if not specified otherwise) with a pre-specified probability. Therefore, economic capital is often calculated as 'value at risk'. In the area of credit risk, economic capital or simply capital, is usually calculated from the quantiles of the loss distribution of an instrument or a portfolio of instruments.

The concept of economic capital differs from regulatory capital in the sense that regulatory capital is the mandatory capital the regulators require to be maintained while economic capital is the best estimate of required capital that financial institutions use internally to manage their own risk and to allocate the cost of maintaining regulatory capital among different units within the organizations.

This thesis will mainly deal with the problem of how to calculate capital for securitizations and re-securitizations. There have been a lot of studies both in academics and in industry on the loss distribution of a credit portfolio and on the loss distribution of a securitization tranche supported by a credit portfolio, since securitizations like RMBS have existed in the U.S. since 1970. However, a lot of re-securitization deals came onto the scene only after 2000. They are more complicated deals than securitizations, and they are quite different from securitizations in a few very important aspects. Regrettably, a lot of people in the industry did not realize or put enough emphasis on those differences, which leads to them buying a lot of re-securitization products (e.g., CDO tranches) without fully understanding the risk that accompanies these products.

What compounded the problem was, at least as it appeared throughout the crisis, that the rating agencies that rate these products (e.g., S&P, Moody's, Fitch) did not model the re-securitization products accurately enough. However, a lot of investors were very much dependent on these agency ratings. Since they do not analyze the risk profile of the products themselves, they depend on the ratings to judge the riskiness of these products. A lot of them were under the 'illusion' that a corporate bond, an RMBS tranche and a CDO tranche have exactly the same risk if they have the same ratings. Of course, as it turned out, this is totally a misconception if not more. See, for example, Hull and White (2010).

On the other hand, the Basel II accord, which most regulatory capital requirements are based upon in the U.S. and Europe, did not differentiate capital requirements of securitizations from those of re-securitizations until recently. See Basel Committee on Banking Supervision (2009) for more details. Additionally, the capital requirements in Basel II are largely tied to agency ratings as well. Retrospectively, the inadequate amount of capital kept by some financial institutions contributed to their downfall as well.

2.1.3 Valuation vs. Credit Models

The typical 'wall-street' approach toward the structured products has been by copula-based simulations, which could capture to some degree the correlation between different obligors and the complicated priority of payments (also called 'waterfall') in those securitization products. This approach is relatively accurate because of the incorporation of deal by deal specifics, but as most Monte Carlo simulation methods it is rather time-consuming. These products were very actively traded before the crisis, so the main objective of the people running this type of model is to get an accurate valuation or pricing. The simulation approach is appropriate for this purpose to calibrate the model to market prices down to nickels and dimes. Refer to Li (2000) and Hull and White (2010) for discussions of copula-based simulation models.

However, for the purpose of setting capital requirements of hundreds or thousands of securitization and re-securitization tranches that a large bank or insurance company hold, the simulation approach is not ideal because of its time-consumingness. One is more concerned about the possible extreme credit losses on the portfolio than whether the portfolio is worth 95 cents on a dollar or 96 cents. It is even a heavier computational burden if one needs to get an accurate estimate of a high quantile of the loss distribution. That is why a credit model with a simplified, fast and analytical or at least half analytical approach would be more desirable. Such a methodology would help setting and benchmarking regulatory capital standards as well.

Of course, an analytical model will not be able to incorporate details of each deal's waterfall, or other factors such as prepayment risk, or interest rate term structure, but with loss as the main focus, the strengths of such a model still outweighs the weaknesses.

The above facts clearly suggest the need for an analytical model that describes the loss distribution of a re-securitization tranche and from which people can get quantiles rather efficiently. We already have something to build upon, which is the Pykhtin-Dev model (Pykhtin & Dev, 2002b) in the Asymptotic Single Risk Factor

(ASRF) framework.

ASRF is a large array of models defined by Gordy (2003), which has desirable properties in terms of portfolio loss distribution. Pykhtin and Dev (2002b) proposed a specific model in this category that has close ties with the Gaussian Copula model in order to calculate the loss distribution and capital of securitizations.

In the first part of this chapter, we will extend the Pykhtin-Dev model to re-securitizations and show how to calculate analytically the loss distribution and capital of a re-securitization tranche. In the second part, we will first generalize the correlation structure in Pykhtin-Dev Model from Gaussian Copula to Student's t Copula for the purpose of incorporating higher tail risk and tail dependence, and then propose a semi-analytical approach to calculate the loss distribution and capital of a securitization tranche in the new model. Both of the new models will be in the ASRF framework.

2.2 Literature Review

2.2.1 Asymptotic Single Risk Factor Models

In his 2003 foundational paper, Gordy defined a category of portfolio credit risk models: the Asymptotic Single Risk Factor (ASRF) model. These models are simple yet powerful tools for analyzing credit risk in portfolios and securitizations. A number of popular credit risk models belong to this category, e.g. the Pykhtin-

Dev model proposed in Pykhtin and Dev (2002b), which the Basel II rating-based approach of capital requirement is based on.

The ASRF model is attractive mainly because of its very desirable asymptotic properties. Some of the important properties are cited below from Gordy (2003) to help develop results in the following sections.

For a portfolio of n obligors, define the portfolio loss ratio L_n as the ratio of total losses to total portfolio exposure, i.e.,

$$L_n \equiv \frac{\sum\limits_{i=1}^{n} U_i A_i}{\sum\limits_{i=1}^{n} A_i}, \tag{2.1}$$

where A_i is the exposure to obligor i, which is known and non-stochastic; U_i is the random variable that denotes loss per dollar exposure. In the event of survival, $U_i = 0$. Otherwise, U_i is the percentage loss-given-default (LGD) on instrument i. For a given $q \in (0, 1)$, value-at-risk (VaR) is defined as the qth quantile of the loss distribution, and is denoted $\mathrm{VaR}_q[L_n]$. Let $\alpha_q(Y)$ denote the qth quantile of the distribution of random variable Y, hence, $\mathrm{VaR}_q[L_n] = \alpha_q(L_n)$.

Let X denote the systematic risk factors of the portfolio, which are drawn from a known joint distribution. It is assumed that all dependence across credit events is due to common sensitivity to these factors. Conditional on X, the remaining credit risk of the portfolio is idiosyncratic to the individual obligors in the portfolio.

In general, X can be multi-dimensional or uni-dimensional. Multi-dimensional X probably makes more intuitive sense because one can usually think of multiple dimensions of systematic factors in a (mortgage) portfolio. Geography, maturity, and fixed rate vs. adjustable rate are all possible dimensions. However, a unidimensional systematic factor offers a great advantage in calculating the loss distribution and capital, as we will see later. Even very simplistic multi risk factor models are usually reliant on computationally intensive simulation methods, while single risk factor models can have elegant analytical solutions. There are methodologies that approximate a multi-risk factor model by a single risk factor model (and an adjustment term), which in essence resembles the principal component method. Please refer to Pykhtin (2004) for an introduction to multi-risk factor models. In this thesis, we will focus on single risk factor models.

Formally, we will assume that

($\mathfrak{A}$-1) the $\{U_i\}$ are bounded in the interval $[-1, 1]$ and, conditional on X, are mutually independent.

($\mathfrak{A}$-2) the A_i are a sequence of positive constants such that

(a) $\sum_{i=1}^{n} A_i \uparrow \infty$ and

(b) there exists a $\zeta > 0$ such that $A_n / \sum_{i=1}^{n} A_i = O(n^{-(1/2+\zeta)})$.

The restrictions in ($\mathfrak{A}$-2) are sufficient to guarantee that the share of the largest single exposure in total portfolio exposure vanishes to zero as the number of exposures in the portfolio increases. As a practical matter, the restrictions are quite weak and would be satisfied by any conceivable real-world portfolio. Gordy (2003) proved the following theorem.

Theorem 1. If ($\mathfrak{A}$-1) and ($\mathfrak{A}$-2) hold, then, conditional on $X = x$, $L_n - E[L_n|x] \to 0$, almost surely.

In intuitive terms, Theorem 1 says that as the exposure share of each asset in the portfolio goes to zero, idiosyncratic risk in portfolio loss is diversified away perfectly. In the limit, the loss ratio converges to a fixed function of the systematic factor X. We refer to this limiting portfolio as "infinitely fine-grained" or as an "asymptotic portfolio." An implication is that, in the limit, we only need to know the conditional distribution of $E[L_n|X]$ to answer any questions about the unconditional distribution of L_n.

Let F_n denote the cumulative distribution function (cdf) of L_n. The second theorem proved by Gordy (2003) can be described as follows.

Theorem 2. If ($\mathfrak{A}$-1) and ($\mathfrak{A}$-2) hold, then for any $\epsilon > 0$,

$$F_n\big(\alpha_q(E[L_n|X]) + \epsilon\big) \to [q, 1], F_n\big(\alpha_q(E[L_n|X]) - \epsilon\big) \to [0, q]. \qquad (2.2)$$

The importance of Theorem 2 is that it allows us to substitue the quantiles of $E[L_n|X]$ (which often times are easier to calculate) for the corresponding quantiles of the loss ratio L_n (which are usually difficult to calculate) as the portfolio becomes large.

Furthermore, the quantiles of $E[L_n|X]$ take on a particularly simple and desirable asymptotic form when we impose two additional restrictions:

($\mathfrak{A}$-3) the systematic risk factor X is one-dimensional; and

($\mathfrak{A}$-4) $E[U_i|x]$ are nondecreasing in x for all i.

If we define

$$M_n(x) \equiv E[L_n|x] = \frac{\sum_{i=1}^{n} E[U_i|x]A_i}{\sum_{i=1}^{n} A_i}, \tag{2.3}$$

the third theorem proven in Gordy (2003) is stated as follows.

Theorem 3. If ($\mathfrak{A}$-3) and ($\mathfrak{A}$-4) are satisfied, then $\alpha_q(E[L_n|X]) = E[L_n|\alpha_q(X)] = M_n(\alpha_q(X))$.

The importance of this result lies in the linearity of the expectation operator. Whereas $\alpha_q(E[L_n|X])$ may, in the general case, be highly complicated, $E[L_n|\alpha_q(X)]$ is simply the exposure-weighted average of the individual assets' conditional expected losses. Taken together with Theorems 1 and 2, Theorem 3 permits a simple and powerful rule for determining capital requirements. For asset i, allocate capital per dollar book value (inclusive of expected loss) of $c_i \equiv E[U_i|\alpha_q(X)] + \epsilon$, for some arbitrarily small ϵ. Observe that this capital charge depends only on the characteristics of instrument i and thus this rule is portfolio-invariant.

Models that satisfy all the assumptions ($\mathfrak{A}$-1) to ($\mathfrak{A}$-4), thus having the properties in Theorems 1, 2 and 3, are usually referred to as the Asymptotic Single Risk

Factor (ASRF) model.

Please refer to Gordy (2003) for proofs and details of the Theorems.

2.2.2 Pykhtin-Dev Model

Pykhtin and Dev (2002a) developed an analytical model to calculate the loss distribution and economic capital of securitizations under the ASRF framework. Because of the close connection of their model to our main model of this chapter, The detailed discussion of the Pykhtin-Dev model will be in Section 2.3.3.1.

2.3 Credit Risk Model for Re-securitizations

2.3.1 Introduction

In the financial turmoil that started in the latter half of 2007, clearly a disproportionate role has been played by the billions of dollars of marked-to-market losses suffered by financial institutions on Collateralized Debt Obligations (CDOs) created out of tranches of Residential Mortgage-backed Securities (RMBS), which in turn were created from subprime and Alt-A mortgages. Such CDOs may be cash CDOs or synthetic CDOs (Credit Default Swaps written with a super-senior tranche of a CDO of RMBS as reference entity).

CDO of RMBS is a special case of recent types of structured credit products known as 're-securitization'. The average investor has always been cognizant of the fact that a re-securitization like a CDO of RMBS is more complex and less transparent than a securitization like RMBS. Yet the average investor has also been content with the rating. After all, both the senior-most tranches of a CDO of RMBS and of a RMBS are rated AAA.

But it is not enough to talk about complexity and transparency in order to have a good understanding of the relative risks of an investment in AAA CDO of RMBS tranche and of an investment in AAA RMBS tranche. The fact that both of them are AAA-rated seems to indicate, at a first glance, that the credit risk in the two types of securities must be rather similar; especially if the structures (securitization and re-securitization) are created out of the same pool of underlying mortgages. In this paper, we show that this statement is not necessarily true.

It has progressively become clear over time that there has been rather an almost complete lack of understanding of the credit risk in re-securitizations such as CDO of RMBS, particularly tail risk measures. The Basel Committee has recently published a proposal revamping the capital adequacy rules for securitization. The proposal contains different sets of capital factors for securitization and re-securitization (*BCBS*, 2009), making a clear distinction between securitization and re-securitization.

Yet there is very little in the published credit risk literature that models credit risk in re-securitization explicitly and that is distinct from credit risk in securitization. All the improvements suggested in the ratings of structured credit products by the rating agencies have been mostly ad-hoc and in any case do not address this basic distinction.

In this section we develop a complete model for credit risk in re-securitizations. The model applies to all re-securitizations which are created from primitive assets that are highly diversified. However, for presentation conciseness, we focus specifically on mortgage-related re-securitizations namely, CDO of RMBS.

The main objective of this section is to present a model for credit risk in re-securitizations. The numerical results are for illustration only. The numerical results from the model show that, starting from the same pool of mortgages, a super-senior RMBS tranche rated AAA and a super-senior CDO tranche rated AAA actually have very different credit risk characteristics.

2.3.2 Some Qualitative Observations

We can make several qualitative observations on the difference between a securitization tranche and a re-securitization tranche.

The fundamental difference lies in the fact that they have collateral support that have very different characteristics. For securitization tranches, usually the collateral consists of stand-alone bonds, loans, or in the case of RMBS, mortgages; while for re-securitization tranches, the collateral consists of RMBS tranches or tranches of other Asset-Backed Securities (ABS).

Mortgages and RMBS tranches have three big differences: first, in terms of loss-given-default, a mortgage rarely has a 100% loss when it defaults. Even in a hostile environment like recently when housing prices have decreased a lot, LGD on a mortgage is usually between 50% and 60% at worst. Other asset-backed securities share similar collateral LGD levels as well. For example, Table 2.1 shows the Moody's historical recovery rates on corporate bonds. Recovery rate is defined as the complement of loss-given-default. One can see that senior secured bonds, which have the highest priority among all corporate bonds, exhibit an average recovery rate of above 50%. Meanwhile, the unsecured or subordinate bonds have average recovery around 30%. However, thin RMBS tranches tend to have very high loss levels when they default, which conceptually can be understood easily, since once the tranche is hit, it only takes a few more defaults before it is completely gone. It turns out that the mezzanine RMBS tranches that people use most to construct CDO deals are usually very thin (around 1%), which leads to very high LGD's.

For the same reason above, one can also say that the loss distributions of a mortgage and an RMBS tranche are very different. The support of the loss distribution of a mortgage rarely includes the 100% point. However, the loss distribution of an mezzanine RMBS tranche will have point masses on the 0% point and the 100% point, since it could incur no loss at all or it could lose every penny of its principal.

Secondly, individual mortgages have relatively low correlation between each other, because of high idiosyncracy. The Basel II standard for correlation between mortgages is 15%. Conversely, RMBS tranches have rather high correlation

Table 2.1: Moody's recovery rates on corporate bonds, 1982-2003.

Class	Average recovery rate (%)
Senior secured	51.6
Senior unsecured	36.1
Senior subordinated	32.5
Subordinated	31.1
Junior subordinated	24.5

Note. The recovery rates are expressed as a percentage of face value.

between each other, for the fact that through the securitization process, idiosyncracies among different mortgages are largely diversified away. As Gordy's results have shown, the risk left is mainly systematic risk. Although different RMBS tranches do not necessarily share the same systematic risk, but they will tend to be much more highly correlated.

The third point is relatively subtle. There have been quite some debates on whether defaults and LGD's are correlated for (corporate) bonds, with the majority of people leaning toward the positive answer. Efforts have been made to incorporate such correlations in the determination of capital, e.g., by the introduction of 'Downturn LGD'. Refer to, for example, Miu and Ozdemir (2006) for a detailed discussion on Downturn LGD. This type of correlation also applies to the case of RMBS tranches, but in a more evident way. For instance, when housing prices decline, not only will the defaults on RMBS tranches go up, but the LGD on RMBS tranches will also elevate, which is exactly what we have been observing in the current crisis.

Higher LGD and higher correlation as stated above make the loss distribution of the collateral portfolio of an RMBS and a CDO very different. As a result, the loss distribution of an RMBS tranche and a CDO tranche will be very different. The proposed new model will naturally incorporate all these considerations.

On the other hand, a lot of financial institutions and rating agencies did not recognize the difference between a regular bond and a securitization tranche, which probably is one of the main reasons why their models failed in this crisis.

2.3.3 The Model

Our objective is to develop a model for credit risk in re-securitization (in particular CDO of RMBS). In doing so, we adopt a "look-through" approach which means we start from the credit characteristics of the underlying primitive loans. In that spirit we first introduce the basic notations for credit risk metrics for the portfolio of loans (in particular mortgages) and for the securitization tranches (in particular RMBS) underneath the re-securitization in subsection 2.3.3.1, where we make use of already existing results in the literature. Discussion of our model really starts from section 2.3.3.2.

2.3.3.1 Loss Distribution of RMBS Tranche - Securitization

Consider an RMBS which has underlying collateral of M homogeneous mortgages with the same probability of default (PD) p, expected loss-given-default (LGD) μ, asset-value correlation (AVC) ρ and equal weights. We will focus on homogeneous portfolios to keep equations and derivations simple. As pointed out later, our results can be extended to non-homogeneous portfolios in a rather straightforward way.

Following Pykhtin and Dev (2002a), we have a set of continuous random variables $\{V_j\}_{j=1}^{M}$ that describe the financial well being of each mortgage. These random variables have standard normal distribution and trigger defaults whenever $V_j < V_j^D \equiv N^{-1}(p)$. $N(\cdot)$ and $N^{-1}(\cdot)$ stand for the cumulative distribution function (CDF) and the inverse cumulative distribution function of the standard normal distribution, respectively. They can be written as

$$V_j = \sqrt{\rho}Y + \sqrt{1-\rho}W_j, \tag{2.4}$$

where Y is the systematic risk factor and $\{W_j\}$ are idiosyncratic factors. All these factors are independent of each other and have standard normal distribution. We also assume independent (independent of Y and $\{W_j\}$ as well) and identically distributed LGD's with expected value μ. In general, $\{V_j\}$ could stand for the asset value of a firm, or the value of a bond, etc. Then the above equation describes that a bond defaults when the firm's asset value drops under a certain threshold.

So this is a very general model that covers other securitizations as well, such as Collateralized Bond Obligations (CBO), Collateralized Loan Obligations (CLO) and Asset-Backed Securities (ABS).

Please note that the systematic risk factor in this thesis here and afterwards is negatively correlated to the loss of the mortgage, as one can easily see from equation (2.4). This is the most intuitive and widely accepted way of specifying loss and asset value. So exactly opposite to ($\mathfrak{A}$-4), $E[U_i|x]$ will be nonincreasing in x for all i instead of nondecreasing. But the quantile argument of Theorem 3 still holds, since one just needs to change the sign of the systematic factor to apply the result. Or instead of

$$\alpha_q(E[L_n|X]) = M_n(\alpha_q(X)), \tag{2.5}$$

one will have this equation:

$$\alpha_q(E[L_n|X]) = M_n(\alpha_{1-q}(X)). \tag{2.6}$$

We will make use of this slightly altered Gordy's result multiple times from now. However, we will not mention this change of sign every time to avoid repetition.

Suppose M is large enough so that the portfolio can be considered fine-grained. Because losses in all the mortgages are affected by a single systematic risk factor, this fits right into the ASRF framework. Thus the loss distribution of the collateral can be approximated by the limiting loss or expected loss given the systematic factor Y:

$$\begin{aligned} L^{\infty}(Y) &\equiv E\left[L^{\text{Coll}}|Y\right] = \sum_{j=1}^{M} E\left[LGD_j \cdot 1_{\{V_j<N^{-1}(p)\}}\middle| Y\right] \\ &= \mu \sum_{j=1}^{M} P\left(\sqrt{1-\rho}W_j < N^{-1}(p) - \sqrt{\rho}Y\right) \qquad (2.7) \\ &= \mu N\left(\frac{N^{-1}(p)-\sqrt{\rho}Y}{\sqrt{1-\rho}}\right), \end{aligned}$$

where LGD_j is the LGD on the jth mortgage and $1_{\{\cdot\}}$ is the indicator function. Thus we have the qth quantile of loss approximately to be $L^{\infty}(Y_{1-q})$ where $Y_{1-q} = N^{-1}(1-q)$.

We denote by $G(l)$ the probability of loss exceeding l, and by (2.7),

$$G(l) = N\left(\frac{1}{\sqrt{\rho}}\left[N^{-1}(p) - \sqrt{1-\rho}N^{-1}\left(\frac{l}{\mu}\right)\right]\right) \quad \text{for} \quad l < \mu. \tag{2.8}$$

Theorem 4 (Pykhtin & Dev, 2002b)

For any tranche $[t_1, t_2]$ such that $t_2 < \mu$, the expected loss (EL) can be computed by

$$E\left[L^{\text{RMBS}}\right] = \frac{\mu}{t_2 - t_1} N_2\left(N^{-1}(p), N^{-1}\left(\frac{t}{\mu}\right); \sqrt{1-\rho}\right)\Bigg|_{t_1}^{t_2}, \tag{2.9}$$

where $N_2(a, b; \gamma)$ denotes the bivariate normal distribution valued at (a, b) with mean 0, standard deviation 1 and correlation γ.

For those tranches $[t_1, t_2]$ such that $t_1 < \mu < t_2$, equation (2.9) will still hold if we replace the term $N^{-1}(t/\mu)$ by $N^{-1}(\min(t/\mu, 1))$. In this case,

$$N_2\left(N^{-1}(p), \infty; \sqrt{1-\rho}\right) = p$$

for any $t \geq \mu$.

Since the portfolio or the tranche loss distribution can be characterized by the single systematic factor Y, we also have the following property:

Corollary 1. The qth quantile of the tranche loss or loss at the confidence level (LCL) is

$$\begin{cases} \dfrac{L^{\infty}\left(Y_{1-q}\right) - t_1}{t_2 - t_1} & \text{if} \quad t_1 < L^{\infty}\left(Y_{1-q}\right) < t_2 \\ 0 & \text{if} \quad L^{\infty}\left(Y_{1-q}\right) < t_1 \\ 1 & \text{otherwise} \end{cases} .$$

One could use this LCL as the capital for the RMBS tranche if looking at the tranche in a stand-alone fashion. But in reality, investors or banks usually buy this tranche and put it into their own investment portfolio. This portfolio is usually very big especially for big banks, and we will refer to this portfolio as the 'super-portfolio' in the future to differentiate it from the collateral portfolio. This concept originates from Pykhtin and Dev (2002b) and is exactly the concept of marginal VaR in the value-at-risk framework (see, for example, Jorion, 2006). So the real question of interest is how much capital the bank should allocate when it adds this RMBS tranche to its super-portfolio. This kind of capital in the context of a super-portfolio will be our focus here and afterwards.

The concept of capital in the context of a super-portfolio not only is more realistic, but also has the advantage of being more smooth than the 'stand-alone' capital (see Pykhtin & Dev, 2002b). As we will see later in the numerical examples, the stand-alone capital tends to be 100% for junior tranches and 0% for senior tranches. This phenomenon makes the stand-alone capital very unsatisfactory particularly for regulatory purpose. (For example, this would mean that literally all 'AAA'-rated tranches will be allowed 'zero' capital.)

We would like to apply Gordy's results in the next step. In order to do that, we need to make some assumptions on the super-porfolio, namely ($\mathfrak{A}$-1)−($\mathfrak{A}$-4). In more practical terms, we have assumed that

i. the super-portfolio where the tranche is held is asymptotically fine-grained;

ii. the super-portfolio is driven by another single systematic risk factor Z; and

iii. investors exposure to the tranche is small compared to total exposure in the super-portfolio.

These are weak enough conditions that most portfolios of big institutions will satisfy. Refer to Pykhtin and Dev (2002b) for details.

Suppose that we place the RMBS tranche into a super-portfolio of which the systematic factor Z is correlated with Y as

$$Y = \sqrt{\lambda} Z + \sqrt{1-\lambda}\eta, \tag{2.10}$$

where the correlation is denoted by λ and η represents a standard normal random variable independent of Z. Additionally, if we assume the super-portfolio satisfies all the requirements of the ASRF model, then in order to calculate the capital for the RMBS tranche in the context of the super-portfolio, we only need to apply Theorem 3 and compute $E\left[L^{\text{RMBS}}|Z\right]$.

First we could compute the probability of loss exceeding level l given the super-portfolio systematic factor Z:

$$G\left(l|Z\right) = N\left(\frac{1}{\sqrt{\rho\left(1-\lambda\right)}}\left[N^{-1}(p) - \sqrt{1-\rho}N^{-1}\left(\frac{l}{\mu}\right) - \sqrt{\rho\lambda}Z\right]\right) \quad \text{for} \quad l < \mu. \tag{2.11}$$

Hence for any tranche $t_2 < \mu$, the conditional expected loss of the tranche will be

$$\begin{aligned} E\left[L^{\text{RMBS}}|Z\right] &= \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} dlG(l|Z) \\ &= \frac{\mu}{t_2 - t_1} N_2\left(\frac{N^{-1}(p) - \sqrt{\rho\lambda}Z}{\sqrt{1-\rho\lambda}}, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\frac{1-\rho}{1-\rho\lambda}}\right)\Bigg|_{t_1}^{t_2}. \end{aligned} \tag{2.12}$$

For those tranches $[t_1, t_2]$ such that $t_1 < \mu < t_2$, we could replace the term $N^{-1}\left(t/\mu\right)$ by $N^{-1}\left(\min\left(t/\mu, 1\right)\right)$ so that (2.12) still holds. Note that for any $t \geq \mu$,

$$N_2\left(\frac{N^{-1}(p) - \sqrt{\rho\lambda}Z}{\sqrt{1-\rho\lambda}}, \infty; \sqrt{\frac{1-\rho}{1-\rho\lambda}}\right) = N\left(\frac{N^{-1}(p) - \sqrt{\rho\lambda}Z}{\sqrt{1-\rho\lambda}}\right).$$

Theorem 5 (Pykhtin & Dev, 2002b)

The capital of an RMBS tranche $[t_1, t_2]$ at the $(100 \times q)\%$ confidence level is simply $E\left[L^{\text{RMBS}}|Z_{1-q}\right]$ as defined in (2.12).

From the above expressions of EL and LCL of an RMBS tranche given in (2.9) and (2.12), respectively, some properties of the two can be seen:

First, for the loss of a whole portfolio, we know that the expected loss does not depend on anything but PD and LGD. But for the loss of an RMBS tranche, not only does it depend on these two parameters, it also depends on the asset value correlation as well. In general, equity tranches' EL decreases when AVC increases, while senior tranches's EL increases when AVC increases. In other words, higher correlation favors equity tranches, but disfavors senior tranches. There is a grey area of mezzanine tranches in the middle, whose direction of movement will depend on their position and other deal specifics.

Second, for LCL or capital in the context of a super-portfolio, correlation to the super-portfolio plays an important role. It is not very straightforward though to see how it impacts the capital. Let's consider extreme situations first:

If $\lambda = 0$, it means that the loss of this tranche is independent of the total performance of the super-portfolio. Plugging $\lambda = 0$ into equation (2.12), we get

$$E\left[L^{\text{RMBS}}|Z\right] = \frac{\mu}{t_2 - t_1} N_2 \left(N^{-1}(p), N^{-1}\left(\frac{t}{\mu}\right); \sqrt{1-\rho} \right) \Bigg|_{t_1}^{t_2} = E\left[L^{\text{RMBS}}\right].$$

So in this case, LCL will be equal to EL for all tranches, meaning that no matter how the super-portfolio performs, the conditional expected loss of the RMBS tranche will always be equal to the unconditional EL.

If $\lambda = 1$, it means that the loss of this tranche is perfectly correlated with the total performance of the super-portfolio. Plug $\lambda = 1$ into equation (2.12) and we get

$$
\begin{aligned}
E\left[L^{\mathrm{RMBS}}|Z\right] &= \frac{\mu}{t_2-t_1} N_2\left(\frac{N^{-1}(p)-\sqrt{\rho}Z}{\sqrt{1-\rho}}, N^{-1}\left(\frac{t}{\mu}\right);1\right)\Bigg|_{t_1}^{t_2} \\
&= \begin{cases} \dfrac{L^{\infty}\left(Y_{1-q}\right)-t_1}{t_2-t_1} & \text{if } \ t_1 < L^{\infty}\left(Y_{1-q}\right) < t_2 \\ 0 & \text{if } \ L^{\infty}\left(Y_{1-q}\right) < t_1 \\ 1 & \text{otherwise} \end{cases},
\end{aligned}
$$

which means that LCL in the context of the super-portfolio will be equal to the stand-alone capital.

In practice, because of its interpretation as the correlation between two systematic factors, λ is usually set between 0.5 and 1. In this case, compared to the stand-alone capital, λ has the effect of transferring capital from junior tranches to senior tranches. Refer to Pykhtin and Dev (2002b) for more details.

2.3.3.2 Loss Distribution of CDO Tranche - Re-securitization

Now we consider a CDO comprised of K homogeneous RMBS mezzanine tranches $[t_1, t_2]$ taken from different mortgage pools. Suppose the loss of the collateral (i.e., the underlying RMBS mezzanine tranches) is driven by a single common factor X, which follows a standard normal distribution and is correlated with each Y_i by ρ_1 as:

$$
Y_i = \sqrt{\rho_1}X + \sqrt{1-\rho_1}\xi_i \tag{2.13}
$$

and ξ_i for $i = 1$ to K represents the idiosyncratic term which follows a standard normal distribution and is independent of X.

Here we will assume additionally that $\mu \geq t_2$, which means the detachment point of the RMBS tranche is below the expected LGD of mortgages. This is nearly a trivial assumption since it is almost universally true for all mezzanine tranches. The expressions below will not be as simple if they are to be applied to senior tranches where the assumption above does not hold, but the derivation will remain the same.

The CDO portfolio that the RMBS tranche is put into can be regarded as a "super-portfolio" as we defined previously. If we look at the default probability and expected LGD of each RMBS within the super-portfolio, we can use the previous results of equations (2.11) and (2.12):

$$PD^{\mathrm{RMBS}} = G(t_1|X) = N\left(\frac{1}{\sqrt{\rho\,(1-\lambda)}}\left[N^{-1}(p) - \sqrt{1-\rho}N^{-1}\left(\frac{t_1}{\mu}\right) - \sqrt{\rho\lambda}X\right]\right),$$

and

$$\begin{aligned} ELGD^{\mathrm{RMBS}} &= E\left[L^{\mathrm{RMBS}}|X\right]/PD^{\mathrm{RMBS}} \\ &= \frac{\mu}{PD^{\mathrm{RMBS}}(t_2-t_1)} N_2\left(\frac{N^{-1}(p)-\sqrt{\rho\lambda}X}{\sqrt{1-\rho\lambda}}, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\frac{1-\rho}{1-\rho\lambda}}\right)\Bigg|_{t_1}^{t_2}. \end{aligned}$$

Now we can see that the default probability and expected LGD of the RMBS tranches are correlated through the common systematic factor X, although not in a simplistic way. For example, the two are not necessarily monotonic with respect to each other. The relationship will depend on the shape of the loss distribution of underlying mortgages.

Heuristically, one can easily argue that when the thickness of a tranche approaches zero, the LGD of that tranche will approach 100%. This is stated in the following proposition which applies to the unconditional case as well.

Proposition 0 (LGD of an infinitesimally thin tranche)

$$\lim_{t_2 - t_1 \searrow 0} ELGD^{RMBS} = 1.$$

Let $F(s) \equiv E\left[L^{\text{RMBS}} | Z = s; \lambda = \rho_1\right]$ as defined in equation (2.12). If K is large enough, due to diversification, the qth quantile of the limiting loss can be approximated by $F(X_{1-q})$ as indicated by Theorem 3. Similar to the case for RMBS tranches in **Corollary 1**, we can have the following

Proposition 1 (Stand-alone Capital of a CDO tranche)

The qth quantile of tranche loss or loss at the confidence level (LCL) is

$$\begin{cases} \dfrac{F(X_{1-q}) - T_1}{T_2 - T_1} & \text{if } \; T_1 < F(X_{1-q}) < T_2 \\ 0 & \text{if } \; F(X_{1-q}) < T_1 \\ 1 & \text{otherwise} \end{cases} .$$

Because we assumed $\mu \geq t_2$, the probability of loss exceeding l can be written as: $G(l) = N(F^{-1}(l))$. Following this technique, we can derive also the EL of the CDO tranche.

Proposition 2 (Expected Loss of a CDO tranche)

For any CDO tranche $[T_1, T_2]$, the expected loss can be computed by:

$$\begin{aligned} E\left[L^{\mathrm{CDO}}\right] &= \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} dl G(l) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} dl N\left(F^{-1}(l)\right) \\ &= \frac{\mu}{(T_2 - T_1)(t_2 - t_1)} \left[Q(\eta_2) - Q(\eta_1)\right], \end{aligned} \tag{2.14}$$

where

$$\eta_i = \frac{N^{-1}(p) - \sqrt{\rho\rho_1} F^{-1}(T_i)}{\sqrt{1 - \rho\rho_1}} \qquad i = 1, 2$$

$$Q(\eta) \equiv N_3 \left(N^{-1}(p), N^{-1}\left(\frac{t}{\mu}\right), \eta; \Sigma \right) \Bigg|_{t_1}^{t_2},$$

$$\Sigma = \begin{bmatrix} 1 & \sqrt{1-\rho} & \sqrt{1-\rho\rho_1} \\ \sqrt{1-\rho} & 1 & \sqrt{\dfrac{1-\rho}{1-\rho\rho_1}} \\ \sqrt{1-\rho\rho_1} & \sqrt{\dfrac{1-\rho}{1-\rho\rho_1}} & 1 \end{bmatrix}$$

and $N_3(a, b, c; \Sigma)$ denotes the trivariate normal distribution valued at (a, b, c) with mean zero and covariance matrix Σ.

If a bank places the CDO tranche into their investment portfolio, which can be regarded as a super-portfolio with the systematic factor Z, and it is correlated with X as

$$X = \sqrt{\lambda_1} Z + \sqrt{1 - \lambda_1} \zeta$$

in which the correlation is denoted by λ_1 and ζ represents a standard normal random variable independent of Z. Together with (2.10) and (2.13), it implies $\lambda = \lambda_1 \rho_1$.

Under this model assumption, the conditional probability of loss exceeding l given the systematic factor can be written as:

$$G\left(l|Z\right) = N\left(\frac{F^{-1}\left(l\right) - \sqrt{\lambda_1}Z}{\sqrt{1-\lambda_1}}\right).$$

Then the conditional expected loss can be written as:

$$\begin{aligned} E\left[L^{\mathrm{CDO}}|Z\right] &= \frac{1}{T_2-T_1}\int_{T_1}^{T_2} dlG(l|Z) = \frac{1}{T_2-T_1}\int_{T_1}^{T_2} dlN\left(\frac{F^{-1}\left(l\right)-\sqrt{\lambda_1}Z}{\sqrt{1-\lambda_1}}\right) \\ &= \frac{\mu}{\left(T_2-T_1\right)\left(t_2-t_1\right)}\left[\tilde{Q}\left(\eta_2\right)-\tilde{Q}\left(\eta_1\right)\right], \end{aligned} \tag{2.15}$$

where

$$\tilde{Q}\left(\eta\right) \equiv N_3\left(\frac{N^{-1}(p)-\sqrt{\rho\rho_1\lambda_1}Z}{\sqrt{1-\rho\rho_1\lambda_1}}, N^{-1}\left(\frac{t}{\mu}\right), \eta; \tilde{\Sigma}\right)\Bigg|_{t_1}^{t_2}$$

and

$$\tilde{\Sigma} = \begin{bmatrix} 1 & \sqrt{\frac{1-\rho}{1-\rho\rho_1\lambda_1}} & \sqrt{\frac{1-\rho\rho_1}{1-\rho\rho_1\lambda_1}} \\ \sqrt{\frac{1-\rho}{1-\rho\rho_1\lambda_1}} & 1 & \sqrt{\frac{1-\rho}{1-\rho\rho_1}} \\ \sqrt{\frac{1-\rho\rho_1}{1-\rho\rho_1\lambda_1}} & \sqrt{\frac{1-\rho}{1-\rho\rho_1}} & 1 \end{bmatrix}.$$

Thus, we have the following proposition:

Proposition 3 (Capital of a CDO tranche in the context of a super-portfolio)

The capital of the CDO tranche $[T_1, T_2]$ within the super-portfolio can be calculated as $E\left[L^{\mathrm{CDO}}|Z_{1-q}\right] - E\left[L^{\mathrm{CDO}}\right]$ as defined in (2.14) and (2.15).

We can see that in the re-securitization model, as we intended, LGDs of RMBS tranches are now derived from their underlying collateral characteristics rather than directly specified as in LGDs of mortgages in the securitization model. Correlation between the RMBS tranches ρ_1 is also an additional parameter that we can set to be much higher than AVC between mortgages.

What is still similar to the securitization model is that ρ_1 now has the same effect that ρ had before. Higher correlation between RMBS tranches will favor low CDO tranches and disfavor high CDO tranches. The correlation to the super-portfolio still plays a similar role as in the securitization model. We will see more concrete examples in the numerical section and get a better view of the parameters.

Note that our results above can be easily extended to the non-homogeneous case. If we assume that the ith RMBS tranche's PD, LGD, tranche limits and weights in the CDO portfolio are p_i, μ_i, t_{i1}, t_{i2} and w_i, respectively. We have the following proposition:

Proposition 4 (Loss distribution of a non-homogeneous CDO)

If Gordy's conditions ($\mathfrak{A}$-1) to ($\mathfrak{A}$-4) are satisfied, we have the same expected loss and capital expressions as in **Proposition 1,2,3**, except that $F(X)$ has to be rewritten as

$$\begin{aligned} F(X) &= \sum_{i=1}^{K} w_i E\left[L_i^{\mathrm{RMBS}} | X\right] \\ &= \sum_{i=1}^{K} \frac{w_i \mu_i}{t_{i2} - t_{i1}} N_2 \left(\frac{N^{-1}(p_i) - \sqrt{\rho\lambda} X}{\sqrt{1-\rho\lambda}}, N^{-1}\left(\frac{t}{\mu_i}\right); \sqrt{\frac{1-\rho}{1-\rho\lambda}} \right) \Bigg|_{t_{i1}}^{t_{i2}}. \end{aligned}$$

2.3.3.3 Granularity Adjustment

While the number M of underlying mortgages in a RMBS is very large, the number of underlying RMBS tranches K in a CDO of RMBS may not be. Therefore, a granularity adjustment becomes necessary. The loss of a coarse-grained CDO with K RMBS tranches can be written as:

$$L_K^{CDO} = E\left[L^{\text{CDO}}|X\right] + R\left(X\right) + O\left(K^{-2}\right)$$

in which $R\left(X\right)$ stands for the first order adjustment. If we neglect terms of higher orders than K^{-1} and let $F\left(X\right) = E\left[L^{\text{CDO}}|X\right] + R\left(X\right)$, the qth quantile of the CDO portfolio loss can still be approximated by $F(X_{1-q})$, and the probability of loss exceeding l can still be written as $G\left(l\right) = N\left(F^{-1}\left(l\right)\right)$. As a result, the stand-alone capital of a coarse-grained CDO tranche will have a similar expression as in **Proposition 1** with the newly defined $F(X)$.

Accordingly, to derive the expected loss of a coarse-grained CDO tranche, one can follow the integration by parts as in (2.14):

$$\int_{T_1}^{T_2} dlN\left(F^{-1}\left(l\right)\right) = \left[N\left(s\right)F\left(s\right) - \int ds\ n\left(s\right)F\left(s\right)\right]\Bigg|_{F^{-1}(T_1)}^{F^{-1}(T_2)}.$$

The computation will have to be carried out by numerical integration in this case. One important thing to note here is that the above granularity adjustment for EL may not work for equity tranches or some 'junior' mezzanine tranches that are just above the equity tranche in the capital structure. Because when losses are small, or X is not in its left tail, the granularity adjustment term is not valid.

Now the only task left is for us to derive the expression of the granularity adjustment term $R(x)$. Let $\pi(x) \equiv E\left[L^{\mathrm{RMBS}}|X=x\right]$ and $\nu(x) \equiv Var\left[L^{\mathrm{RMBS}}|X=x\right]$. Following Gordy (2003) and Wilde (2001),

$$R(x) = -\frac{1}{2K\pi'(x)}\left[v'(x) - v(x)\left(\frac{\pi''(x)}{\pi'(x)} + x\right)\right]. \tag{2.16}$$

Let

$$\tilde{x} \equiv \frac{N^{-1}(p) - \sqrt{\rho\rho_1}x}{\sqrt{1-\rho\rho_1}} \quad \text{and} \quad \tilde{\rho} \equiv \frac{1-\rho}{1-\rho\rho_1}.$$

Equation (2.12) leads to

$$\pi(x) = \frac{\mu}{t_2 - t_1} N_2\left(\tilde{x}, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\tilde{\rho}}\right)\Bigg|_{t_1}^{t_2}. \tag{2.17}$$

Then we can take the first and second derivatives:

$$\pi'(x) = \frac{\mu}{t_2 - t_1}\left(\sqrt{\frac{\rho\rho_1}{1-\rho\rho_1}}\right) n(\tilde{x})\, N\left(\frac{\sqrt{\tilde{\rho}}\tilde{x} - N^{-1}(t/\mu)}{\sqrt{1-\tilde{\rho}}}\right)\Bigg|_{t_1}^{t_2} \tag{2.18}$$

$$\begin{aligned}\pi''(x) = & \frac{\mu}{t_2 - t_1}\left(\frac{\rho\rho_1}{1-\rho\rho_1}\right) n(\tilde{x}) \\ & \cdot\left[\tilde{x}N\left(\frac{\sqrt{\tilde{\rho}}\tilde{x} - N^{-1}(t/\mu)}{\sqrt{1-\tilde{\rho}}}\right) - \sqrt{\frac{\tilde{\rho}}{1-\tilde{\rho}}}\, n\left(\frac{\sqrt{\tilde{\rho}}\tilde{x} - N^{-1}(t/\mu)}{\sqrt{1-\tilde{\rho}}}\right)\right]\Bigg|_{t_1}^{t_2}\end{aligned} \tag{2.19}$$

in which

$$H(t) \equiv \sqrt{\frac{1}{1-\tilde{\rho}}}N^{-1}\left(\frac{t}{\mu}\right).$$

The conditional loss variance and its derivative can be calculated as follows:

$$\nu(x) = \frac{2\mu^2}{(t_2-t_1)^2} N_3\left(N^{-1}\left(\frac{t}{\mu}\right), 0, \tilde{x}; \hat{\Sigma}\right)\Bigg|_{t_1}^{t_2} - \frac{2t_1}{t_2-t_1}\pi(x) - \pi^2(x) \tag{2.20}$$

in which

$$\hat{\Sigma} = \begin{bmatrix} 1 & -\sqrt{1/2} & \sqrt{\tilde{\rho}} \\ -\sqrt{1/2} & 1 & -\sqrt{\tilde{\rho}/2} \\ \sqrt{\tilde{\rho}} & -\sqrt{\tilde{\rho}/2} & 1 \end{bmatrix},$$

and

$$\begin{aligned} \nu'(x) = \quad & n(\tilde{x})\, N_2 \left(\frac{\sqrt{\tilde{\rho}}\tilde{x} - N^{-1}(t/\mu)}{\sqrt{1-\tilde{\rho}}}, \sqrt{\frac{\tilde{\rho}}{2-\tilde{\rho}}}\tilde{x}; \sqrt{\frac{1-\tilde{\rho}}{2-\tilde{\rho}}} \right) \Bigg|_{t_1}^{t_2} \\ & \cdot \frac{2\mu^2}{(t_2-t_1)^2} \left(\sqrt{\frac{\rho\rho_1}{1-\rho\rho_1}} \right) - \frac{2t_1}{t_2-t_1} \pi'(x) - 2\pi(x)\,\pi'(x)\,. \end{aligned} \tag{2.21}$$

Plugging equations (2.17) - (2.21) into (2.16), we can obtain the granularity adjustment term $R(x)$.

Corollary 2 (Granularity Adjustment term for a coarse-grained CDO)

Using the above notations, the granularity adjustment term for a coarse-grained CDO can be calculated as

$$R(x) = -\frac{1}{2K\pi'(x)} \left[v'(x) - v(x) \left(\frac{\pi''(x)}{\pi'(x)} + x \right) \right],$$

where $\pi'(x)$, $\pi''(x)$, $v(x)$ and $v'(x)$ are as defined in equations (2.17) - (2.21).

2.3.3.4 Numerical Results

The CDO we consider here in this section has five different tranches and is made up of $K = 30$ RMBS mezzanine tranches. The RMBS tranches are homogeneous with tranche limits equal to (3%, 5%), PD = 3%, LGD = 20% and AVC = 0.15.

Confidence level $q = 99.9\%$ is used, and the correlation between the CDO and the super-portfolio is assumed to be 0.9. Results are also shown with a range of possible correlation between RMBS tranches, from 0.5 to 0.8.

For comparison to our theoretical results, we also performed a 'look-through' simulation on the CDO, where loss distributions of RMBS tranches are generated according to equation (2.7) and then aggregated for computation of EL and loss quantile or loss at the confidence level (LCL) of the CDO. All simulation results are based on 1,000,000 Monte Carlo samples.

The following columns are displayed in Table 2.2: LCL(sim) or LCL from simulation, LCL(sa) or standalone LCL, LCL(adj) or LCL with granularity adjustment, LCL(sp) or LCL in the context of a super-portfolio, EL(sim) or EL from simulation, EL(grn) or EL from analytical solution assuming granular CDO portfolio, and EL(adj) or EL with granularity adjustment.

It can be seen from the tables that across different RMBS tranche correlations, the theoretical LCL values are fairly close to their simulation-based counterparts for most of the tranches. After granularity adjustment, the two values agree even more. The same can be said for expected loss except for the equity tranche, where adjusted EL does not work as noted before.

As in the case of securitizations, the LCL values in the context of a super-portfolio exhibit the phenomenon of transferring capital from junior tranches to senior tranches. And the super-senior tranche of the CDO is not allocated zero

Table 2.2: EL and LCL with $K = 30$

Tranche Limits(%)	LCL(sim)	LCL(sa)	LCL(adj)	LCL(sp)	EL(sim)	EL(grn)	EL(adj)
RMBS Tranche Correlation = 0.5							
0.0 - 6.0	100	100	100	99.2367	3.7666	3.9826	4.8402
6.0 - 7.0	100	100	100	94.9089	1.0023	0.7931	0.9916
7.0 - 15.0	100	100	100	74.2016	0.4605	0.3639	0.4551
15.0 - 50.0	17.0357	11.6144	16.8975	9.8743	0.0504	0.0387	0.0494
50.0 - 100.0	0	0	0	0.0017	0.0006	0.0005	0.0007
RMBS Tranche Correlation = 0.6							
0.0 - 6.0	100	100	100	99.4750	3.2761	3.4898	3.9971
6.0 - 7.0	100	100	100	96.9279	1.1016	0.9679	1.1002
7.0 - 15.0	100	100	100	85.0848	0.5826	0.5123	0.5817
15.0 - 50.0	35.3093	30.0065	34.4230	21.9768	0.0944	0.0814	0.0935
50.0 - 100.0	0	0	0	0.0940	0.0036	0.0025	0.0031
RMBS Tranche Correlation = 0.7							
0.0 - 6.0	100	100	100	99.5047	2.7826	2.9278	3.1624
6.0 - 7.0	100	100	100	97.5858	1.1390	1.0555	1.1313
7.0 - 15.0	100	100	100	90.2197	0.6766	0.6282	0.6729
15.0 - 50.0	58.8438	52.9004	56.4758	37.2015	0.1545	0.1397	0.1509
50.0 - 100.0	0	0	0	1.0598	0.0102	0.0088	0.0100
RMBS Tranche Correlation = 0.8							
0.0 - 6.0	100	100	100	99.3644	2.2087	2.3116	2.3580
6.0 - 7.0	100	100	100	97.5508	1.0768	1.0489	1.0825
7.0 - 15.0	100	100	100	92.4092	0.7079	0.6933	0.7157
15.0 - 50.0	87.2661	82.5996	85.2957	52.7641	0.2154	0.2088	0.2168
50.0 - 100.0	0	0	0	5.1389	0.0269	0.0241	0.0258

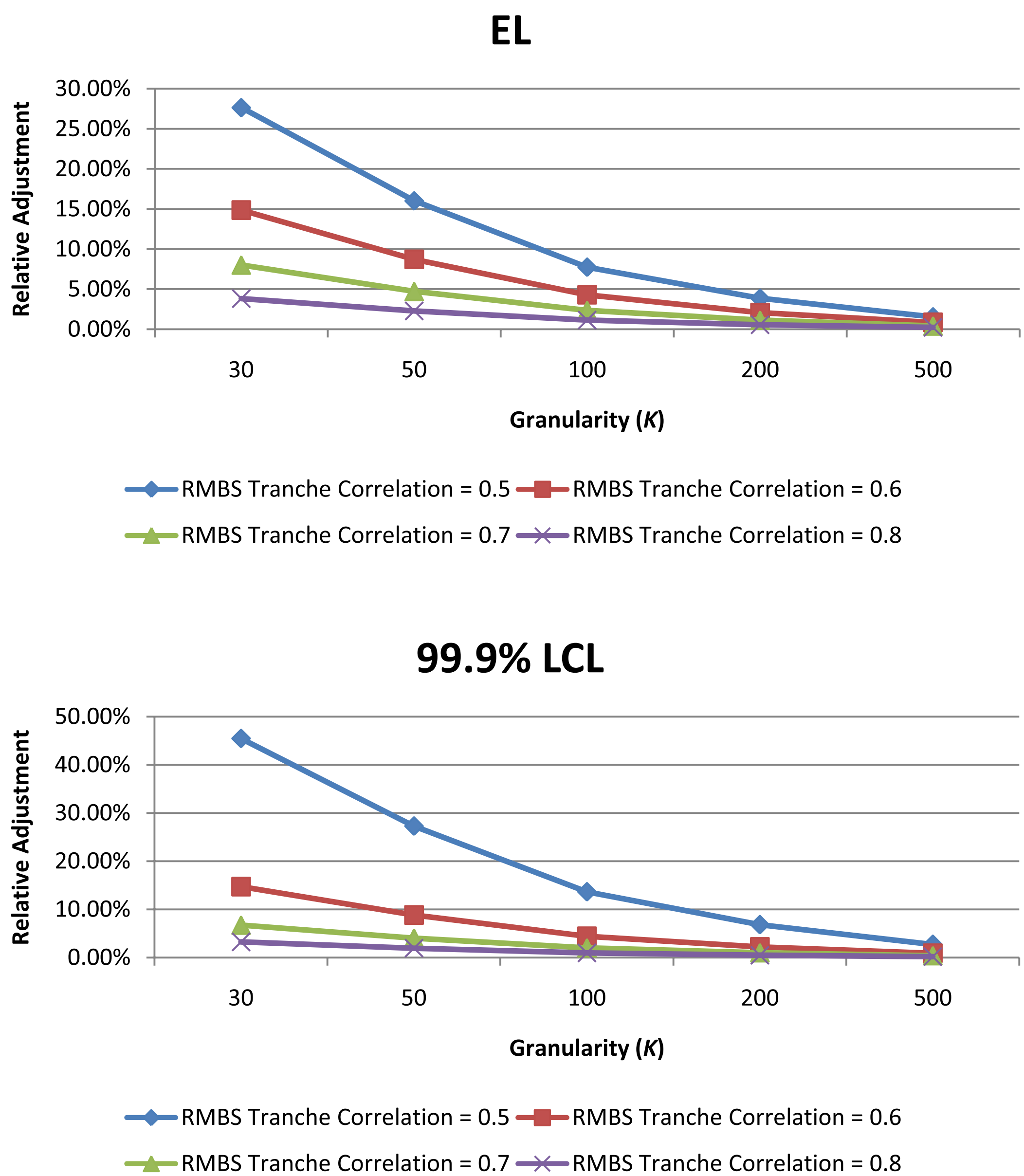
Table 2.3: CDO Tranche 15% - 50%

RMBS Correlation	30	50	100	200	500	∞
			Loss at the 99.9% Confidence Level			
0.5	16.8975	14.7842	13.1993	12.4069	11.9314	11.6144
0.6	34.4230	32.6564	31.3315	30.6690	30.2715	30.0065
0.7	56.4758	55.0456	53.9730	53.4367	53.1149	52.9004
0.8	85.2957	84.2172	83.4084	83.0040	82.7613	82.5996
			Expected Loss			
0.5	0.0494	0.0449	0.0417	0.0402	0.0393	0.0387
0.6	0.0935	0.0885	0.0849	0.0831	0.0821	0.0814
0.7	0.1509	0.1463	0.1430	0.1413	0.1403	0.1397
0.8	0.2168	0.2136	0.2112	0.2100	0.2093	0.2088

capital any more as in other standalone cases.

In Table 2.3, we focus on LCL and EL values for the (15%, 50%) CDO tranche in the previous example with different level of granularity (K). It is straightforward to observe that the granularity adjustment shrinks as K increases and idiosyncratic risks are diversified away. Also noticeable is that the granularity adjustment increases as RMBS tranche correlation decreases, which is also understandable given the fact that lower correlation increases the idiosyncratic risk of a non-granular portfolio. In the extreme case, when RMBS tranche correlation is close to 1, granularity adjustment would not be necessary.

Figure 2.2: Granularity adjustments



2.3.3.5 Conclusion

In securitization, the underlying collateral consists of loans and bonds while in re-securitization the underlying collateral consists of tranches of securitization. In terms of loss given default or equivalently extent of penetration into a tranche and in terms of tail of loss distribution, tranches are nothing like bonds. Hence our premise is that the credit risk in re-securitization is very different from the credit risk in securitization.

In this part we have developed a full-fledged model for credit risk in resecuritization using established principles of portfolio credit risk methodologies. The model is analytically tractable and can, therefore, be quickly implemented. The numerical results illustrate that after applying granularity adjustments, we get essentially the same results as a Monte Carlo simulation.

2.3.4 Application: Are the AAA Ratings Justified?

The term CDO has earned a rather poor reputation in recent memory. In the financial turmoil that started in the latter half of 2007, clearly a disproportionate role has been played by the billions of dollars of marked-to-market losses suffered by financial institutions on CDOs created out of tranches of Residential Mortgage-backed Securities (RMBS), which in turn were created from subprime mortgages. Such CDOs may be cash CDOs or synthetic CDOs (Credit Default Swaps written with a super-senior tranche of a CDO of RMBS as reference entity).

Some writers in the popular press have recently expressed their shock that low quality subprime mortgages could be packaged into securities in effect as creditworthy as U.S. Treasury bonds. Last year, U.S. Congressional testimony of the heads of major U.S. rating agencies indicated that staffers and Congressmen were also appalled at how the rating agencies could have given such pristine ratings to bonds structured out of low quality subprime mortgages. As has been noted: “High ratings given to low-quality assets, particularly those based on risky mortgages, have been criticized by authorities round the world for contributing to the credit market bubbles that have collapsed in the crisis” (Financial Times, October 23, 2008). More recently, the European Commission has come out strongly against the rating agencies as reflected in the following viewpoint: “Agencies gave top ratings to subprime-related structured finance products adding to the crisis of summer 2007. They also later failed to reflect in their ratings the market’s worsening conditions” (Ignites Europe, November 13, 2008). In what follows, we show that the sentiments expressed are not quite logical.

The average investor is typically a ‘AAA’ investor. Most investors and pension funds treat AAA-rated investments in the same category (of safety) as Treasury instruments. The average investor has always been cognizant of the fact that a re-securitization like a CDO of RMBS is more complex and less transparent than a securitization like RMBS. But the fact that both are AAA-rated seems to indicate, at first glance, that the credit risk in the two types of securities must be quite similar, especially if the structures (securitization and re-securitization) are created out of the same pool of underlying subprime mortgages. In what follows, we show that this is far from the case.

We extended the applications of the asymptotic single risk factor (ASRF) model, introduced by Gordy (2003) for credit risk in subprime portfolios, and of the securitization credit risk model introduced in Pykhtin and Dev (2002b). What we will apply are mainly results of this chapter on typical mortgage securitizations and re-securitizations.

A subprime mortgage, by definition, is a loan to a household with poor credit history as reflected in the credit score and consequently the probability of default (PD). In recent years, the steep decline in home prices and the prevalence of poor underwriting documentation on appraisals as well as income have increased the loss given default (LGD) on defaulted subprime mortgages. Prime mortgages, in contrast, are characterized by much lower PD and somewhat lower LGD. The two risk measures we will focus on both for the underlying pool of mortgages and for the tranches of RMBS and the tranches of CDO of RMBS are: Expected Loss (EL) and Unexpected Loss (UL). The former is nothing more than the product of PD and LGD. The latter also gives us an idea of the equity capital that is put at risk to support an investment in the applicable tranche. Mathematically, UL is nothing more than the difference between LCL and EL.

We consider here as a numerical example of CDO senior tranches (with 20%

or 50% credit enhancement underneath) that have RMBS mezzanine and senior tranches as their underlying collateral. Each of the RMBS, in turn, has underlying collateral of thousands of mortgages with homogeneous PD, LGD and asset value correlation (AVC). One low PD (1%) and one high PD (5%) scenarios are studied. UL here is based on the 99.9% confidence level. In order to model re-securitizations, we need one more input variable which is the correlation between different RMBS tranches. Although the range of probable values of this correlation is not clear so far, it has to be significantly greater than AVC between individual mortgages because in an ASRF framework, idiosyncratic risks are diversified away from an RMBS portfolio, leaving the correlation between RMBS tranches to be the correlation between systematic risk factors.

In each sub-table of Table 2.4 and Table 2.5, we show sequentially the EL and UL of the RMBS tranche and the EL and UL of the CDO tranches made out of the corresponding RMBS tranches. The following observations can be made after an examination of the results:

- As pointed out in Pykhtin and Dev (2002b) and acknowledged by most practitioners, thin securitization tranches are generally riskier than thick tranches, mainly because of higher LGDs. As a result, re-securitizations with thin securitization tranches underneath are riskier.

- When PD is relatively low, or for 'near-prime' mortgages, it is possible to

Table 2.4: EL and UL of RMBS and CDO tranches

Low PD (1%) Thick Securitization Tranches				High PD (5%) Thick Securitization Tranches			
Securitization				Securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
3	8	0.0226	3.3129	3	8	3.0210	63.0955
Re-securitization				Re-securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
20	100	0.0017	0.3022	20	100	1.0868	42.1947
50	100	0.0002	0.0258	50	100	0.3348	26.0281
Securitization				Securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
5	10	0.0017	0.3002	5	10	0.6871	37.0500
Re-securitization				Re-securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
20	100	0.0001	0.0110	20	100	0.1882	16.8579
50	100	0.0000	0.0006	50	100	0.0509	7.1212
Securitization				Securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
10	15	0.0000	0.0003	10	15	0.0160	2.6377
Re-securitization				Re-securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
20	100	0.0000	0.0000	20	100	0.0020	0.3494
50	100	0.0000	0.0000	50	100	0.0004	0.0526

Table 2.5: EL and UL of RMBS and CDO tranches (continued)

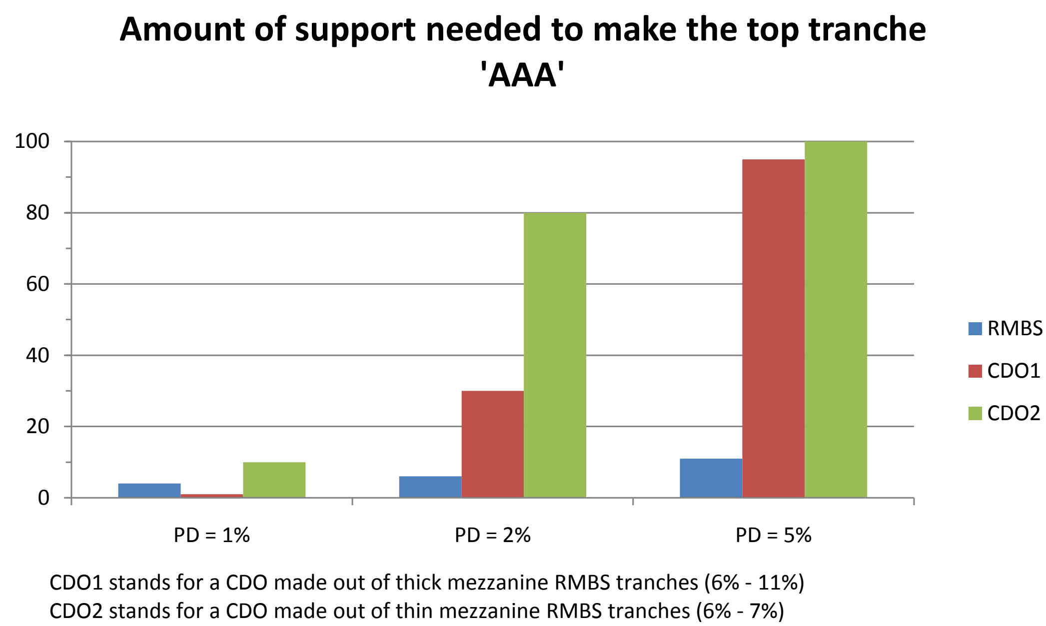
Low PD (1%) Thin Securitization Tranches				High PD (5%) Thin Securitization Tranches			
Securitization				Securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
3	4	0.0832	11.4104	3	4	8.1225	85.6365
Re-securitization				Re-securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
20	100	0.0217	3.4189	20	100	5.0456	75.5386
50	100	0.0066	1.1695	50	100	2.8298	69.4426
Securitization				Securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
5	6	0.0061	1.0913	5	6	1.8361	66.0785
Re-securitization				Re-securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
20	100	0.0010	0.1633	20	100	0.8777	42.8805
50	100	0.0002	0.0307	50	100	0.4000	30.5012
Securitization				Securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
10	11	0.0000	0.0011	10	11	0.0445	6.9995
Re-securitization				Re-securitization			
Attach(%)	Detach(%)	EL(%)	UL(%)	Attach(%)	Detach(%)	EL(%)	UL(%)
20	100	0.0000	0.0000	20	100	0.0108	1.8385
50	100	0.0000	0.0000	50	100	0.0032	0.5647

make high-quality (AAA-rated) CDO tranches with close-to-zero EL out of RMBS mezzanine tranches, although the CDO tranches often have UL that cannot be ignored. On the other hand, for subprime mortgages or in situations where PD rises to 5%, even credit enhancement of 50% is not enough to make the CDO senior tranche safe in terms of EL. Moreover, the corresponding UL is substantial.

- If we compare the senior tranche of a securitization to that of a re-securitization, an RMBS senior tranche with 20% of support has close-to-zero EL and UL, and these two metrics do not suffer big increases when PD elevates to 5%. Conversely, when mortgage quality deteriorates, those seemingly 'safe' CDO senior tranches made out of mezzanine RMBS tranches exhibit huge increases in both EL and UL, and are obviously not 'safe' any more.

The above points could be further explained by Figure 2.3 which shows the amount of support needed to make the top tranche 'AAA' for RMBS and reflects two kinds of CDO. It can be seen that a CDO made out of thin RMBS tranches requires more support than one made out of thick tranches; an RMBS generally requires less support than a CDO constructed from mezzanine RMBS tranches; and the support such a CDO needs to make its most senior tranche 'AAA' increases dramatically as PD increases, to such an extent that when PD is 5%, a AAA rating is virtually impossible to achieve.

Figure 2.3: Amount of support needed to make the top RMBS tranche 'AAA'



In conclusion, it is not irrational to be shocked at how rating agencies could have given such high ratings (AAA) to structures made out of very low credit quality subprime mortgages. A AAA-bond can be created out of a portfolio of subprime mortgages and there is nothing surprising about it. However, whether it is appropriate to do so depends on the securitization structure in question. In this section we show that in a RMBS (securitization) structure, whether made out of a pool of prime or subprime mortgages a AAA tranche can easily be created. Similarly, in a CDO (re-securitization) structure made out of senior RMBS tranches, it is not difficult to create a AAA tranche, with enough credit enhancement; and this can be made out of a pool of prime mortgages or a pool of subprime mortgages. However, in a CDO (re-securitization) structure made out of mezzanine or junior

RMBS tranches with subprime mortgages as underlying collateral, it is almost impossible to create a AAA tranche.

Coincidentally, Hull and White (2010) studied essentially the same problem using Copula and simulation techniques. Although their applied a different methodology, they arrived at essentially the same conclusion as ours:

"Contrary to many of the opinions that have been expressed, the AAA ratings for the senior tranches of ABSs were not unreasonable. The weighted average life of mortgages is about five years. The probability of loss and expected loss of the AAA-rated tranches that were created were similar to or better than those of AAA-rated five-year bonds.

"The AAA ratings for Mezz ABS CDOs are much less defensible. Scenarios where all the underlying BBB tranches lose virtually all their principal are sufficiently probable that it is not reasonable to assign a AAA rating to even a quite thin senior tranche."

2.3.5 Derivation of Results

A. Derivation of RMBS result in Theorem 4

$$\begin{aligned} E\left[L^{\mathrm{RMBS}}\right] &= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} dl G(l) \\ &= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} N\left(\frac{1}{\sqrt{\rho}}\left[N^{-1}(p) - \sqrt{1-\rho} N^{-1}\left(\frac{l}{\mu}\right)\right]\right) dl. \end{aligned}$$

Let $s = N^{-1}\left(\frac{l}{\mu}\right)$. Then

$$
\begin{aligned}
E\left[L^{\mathrm{RMBS}}\right] &= \frac{\mu}{t_2 - t_1}\int_{t_1}^{t_2} N\left(\frac{1}{\sqrt{\rho}}\left[N^{-1}(p) - \sqrt{1-\rho}s\right]\right) dN(s) \\
&= \frac{\mu}{t_2 - t_1}\, N_2\left(N^{-1}(p), N^{-1}\left(\frac{t}{\mu}\right); \sqrt{1-\rho}\right)\Bigg|_{t_1}^{t_2}.
\end{aligned}
$$

B. Derivation of CDO of RMBS results in section 2.3.3.2

$$
\begin{aligned}
E\left[L^{\mathrm{CDO}}\right] &= \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} dlG(l) = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} dlN\left(F^{-1}\left(l\right)\right) \quad (\text{let } s = F^{-1}\left(l\right)) \\
&= \frac{1}{T_2 - T_1}\int_{F^{-1}(T_1)}^{F^{-1}(T_2)} N\left(s\right) dF\left(s\right) \\
&= \frac{\mu}{(T_2 - T_1)(t_2 - t_1)} \cdot \\
&\quad \int_{F^{-1}(T_1)}^{F^{-1}(T_2)} N\left(s\right) dN_2\left(\frac{N^{-1}(p) - \sqrt{\rho\rho_1}s}{\sqrt{1-\rho\rho_1}}, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\frac{1-\rho}{1-\rho\rho_1}}\right)\Bigg|_{t_1}^{t_2}.
\end{aligned}
$$

Let

$$
\eta = \frac{N^{-1}(p) - \sqrt{\rho\rho_1}s}{\sqrt{1-\rho\rho_1}},
$$

and let

$$
\eta_i = \frac{N^{-1}(p) - \sqrt{\rho\rho_1}F^{-1}\left(T_i\right)}{\sqrt{1-\rho\rho_1}}, \quad i = 1, 2.
$$

Then

$$\begin{aligned}
E\left[L^{\mathrm{CDO}}\right] &= \frac{\mu}{(T_2-T_1)(t_2-t_1)} \cdot \\
&\quad \int_{\eta_1}^{\eta_2} N\left(\frac{N^{-1}(p)-\sqrt{1-\rho\rho_1}\eta}{\sqrt{\rho\rho_1}}\right) dN_2\left(\eta, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\frac{1-\rho}{1-\rho\rho_1}}\right)\Bigg|_{t_1}^{t_2} \\
&= \frac{\mu}{(T_2-T_1)(t_2-t_1)} \left[N_3\left(N^{-1}(p), N^{-1}\left(\frac{t}{\mu}\right), \eta; \Sigma\right)\Bigg|_{t_1}^{t_2}\right]\Bigg|_{\eta_1}^{\eta_2} \\
&= \frac{\mu}{(T_2-T_1)(t_2-t_1)} \left[Q(\eta_2) - Q(\eta_1)\right],
\end{aligned}$$

and

$$\begin{aligned}
& E\left[L^{\mathrm{CDO}}|Z\right] \\
= & \frac{1}{T_2-T_1}\int_{T_1}^{T_2} dlG(l|Z) = \frac{1}{T_2-T_1}\int_{T_1}^{T_2} dlN\left(\frac{F^{-1}(l)-\sqrt{\lambda_1}Z}{\sqrt{1-\lambda_1}}\right) \\
= & \frac{1}{T_2-T_1}\int_{F^{-1}(T_1)}^{F^{-1}(T_2)} N\left(\frac{s-\sqrt{\lambda_1}Z}{\sqrt{1-\lambda_1}}\right) dF(s) \\
= & \frac{\mu}{(T_2-T_1)(t_2-t_1)} \cdot \\
& \int_{F^{-1}(T_1)}^{F^{-1}(T_2)} N\left(\frac{s-\sqrt{\lambda_1}Z}{\sqrt{1-\lambda_1}}\right) dN_2\left(\frac{N^{-1}(p)-\sqrt{\rho\rho_1}s}{\sqrt{1-\rho\rho_1}}, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\frac{1-\rho}{1-\rho\rho_1}}\right)\Bigg|_{t_1}^{t_2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mu}{(T_2-T_1)(t_2-t_1)} \cdot \\
&\quad \int_{\eta_1}^{\eta_2} N\left(\frac{N^{-1}(p)-\sqrt{\rho\rho_1\lambda_1}Z-\sqrt{1-\rho\rho_1}\eta}{\sqrt{\rho\rho_1(1-\lambda_1)}}\right) dN_2\left(\eta, N^{-1}\left(\frac{t}{\mu}\right); \sqrt{\frac{1-\rho}{1-\rho\rho_1}}\right)\Bigg|_{t_1}^{t_2} \\
&= \frac{\mu}{(T_2-T_1)\,(t_2-t_1)} \left[N_3\left(\frac{N^{-1}(p)-\sqrt{\rho\rho_1\lambda_1}Z}{\sqrt{1-\rho\rho_1\lambda_1}}, N^{-1}\left(\frac{t}{\mu}\right), \eta; \tilde{\Sigma}\right)\Bigg|_{t_1}^{t_2}\right]\Bigg|_{\eta_1}^{\eta_2} \\
&= \frac{\mu}{(T_2-T_1)\,(t_2-t_1)} \left[\tilde{Q}\,(\eta_2)-\tilde{Q}\,(\eta_1)\right].
\end{aligned}
$$

C. Derivation of granularity adjustment results in section 2.3.3.3

$$
v\,(x) = E\left[L^2|x\right] - \pi^2\,(x)\,.
$$

$$
\begin{aligned}
E\left[L^2|x\right] &= \int_{t_1}^{t_2} \left(\frac{l-t_1}{t_2-t_1}\right)^2 f\,(l|x)\,dl + G\,(t_2) \\
&= \frac{2}{(t_2-t_1)^2}\left[\int_{t_1}^{t_2} G\,(l|x)\cdot l dl - t_1 \int_{t_1}^{t_2} G\,(l|x)\,dl\right] \\
&= \frac{2}{(t_2-t_1)^2}\int_{t_1}^{t_2} G\,(l|x)\cdot l dl - \frac{2t_1}{t_2-t_1}\pi\,(x)\,.
\end{aligned}
$$

$$\begin{aligned}
\int_{t_1}^{t_2} G\left(l|x\right)\cdot l dl &= \int_{t_1}^{t_2} N\left(\frac{N^{-1}(p)-\sqrt{\rho\rho_1}x-\sqrt{1-\rho}N^{-1}\left(l/\mu\right)}{\sqrt{\rho\left(1-\rho_1\right)}}\right)\cdot l dl \\
&= \int_{t_1}^{t_2} N\left(\frac{\tilde{x}-\sqrt{\tilde{\rho}}N^{-1}\left(l/\mu\right)}{\sqrt{1-\tilde{\rho}}}\right)\cdot l dl \qquad \left(\text{let } s=N^{-1}\left(l/\mu\right)\right) \\
&= \mu^2 \int_{N^{-1}(t_1/\mu)}^{N^{-1}(t_2/\mu)} N\left(\frac{\tilde{x}-\sqrt{\tilde{\rho}}s}{\sqrt{1-\tilde{\rho}}}\right) N\left(s\right) n\left(s\right) ds \\
&= \mu^2 \; N_3\left(N^{-1}\left(\frac{t}{\mu}\right),0,\tilde{x};\hat{\Sigma}\right)\Bigg|_{t_1}^{t_2}.
\end{aligned}$$

$$v'\left(x\right)=\frac{2}{\left(t_2-t_1\right)^2}\int_{t_1}^{t_2}\frac{d}{dx}G\left(l|x\right)\cdot l dl-\frac{2t_1}{t_2-t_1}\pi'\left(x\right)-2\pi\left(x\right)\pi'\left(x\right).$$

$$\int_{t_1}^{t_2}\frac{d}{dx}G\left(l|x\right)\cdot l dl=-\sqrt{\frac{\rho\rho_1}{1-\rho\rho_1}}\int_{t_1}^{t_2}\frac{1}{\sqrt{1-\tilde{\rho}}}\;n\left(\frac{\tilde{x}-\sqrt{\tilde{\rho}}N^{-1}\left(l/\mu\right)}{\sqrt{1-\tilde{\rho}}}\right)\cdot l dl.$$

Let $s=N^{-1}\left(l/\mu\right)$. Then

$$\begin{aligned}
\int_{t_1}^{t_2}\frac{d}{dx}G\left(l|x\right)\cdot l dl &= -\mu^2\sqrt{\frac{\rho_1}{1-\rho_1}}\int_{N^{-1}(t_1/\mu)}^{N^{-1}(t_2/\mu)} n\left(\frac{\tilde{x}-\sqrt{\tilde{\rho}}s}{\sqrt{1-\tilde{\rho}}}\right)N\left(s\right)n\left(s\right)ds \\
&= -\mu^2\sqrt{\frac{\rho_1}{1-\rho_1}}\;n\left(\tilde{x}\right)\int_{N^{-1}(t_1/\mu)}^{N^{-1}(t_2/\mu)} n\left(\frac{\tilde{x}-\sqrt{\tilde{\rho}}s}{\sqrt{1-\tilde{\rho}}}\right)N\left(s\right)ds \\
&= \mu^2\sqrt{\frac{\rho\rho_1}{1-\rho\rho_1}}\;n\left(\tilde{x}\right)N_2\left(\frac{\sqrt{\tilde{\rho}}\tilde{x}-N^{-1}\left(t/\mu\right)}{\sqrt{1-\tilde{\rho}}},\sqrt{\frac{\tilde{\rho}}{2-\tilde{\rho}}}\tilde{x};\sqrt{\frac{1-\tilde{\rho}}{2-\tilde{\rho}}}\right)\Bigg|_{t_1}^{t_2}.
\end{aligned}$$

2.4 Modeling Credit Risk in Securitizations Using Student's t Copula

2.4.1 A Review of Copula Functions

In statistics, a copula is used as a general way of formulating a multivariate distribution in such a way that various general types of dependence can be represented. In finance, copula has been widely used in recent year because of the need to model all kinds of dependency between different obligors, bonds or other financial instruments. See for example Li (2000).

Definition 1. An n-dimensional colupa is a function $C : [0,1]^n \to [0,1]$ which has the following properties:

1. $C(\mathbf{u})$ is increasing in each component $\mathbf{u}_k$ with $k \in \{1, 2, \ldots, n\}$.

2. For every vector $\mathbf{u} \in [0,1]^n$, $C(\mathbf{u}) = \mathbf{0}$ if at least one coordinate of the vector $\mathbf{u}$ is 0 and $C(\mathbf{u}) = \mathbf{u}_k$ if all the coordinate of $\mathbf{u}$ are equal to 1 except the k-th one.

3. For every $\mathbf{a}, \mathbf{b} \in [0,1]^n$ with $\mathbf{a} \leq \mathbf{b}$, given a hypercube $\mathbf{B} = [\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times [a_2, b_2] \cdots \times [a_n, b_n]$ whose vertices lie in the domain of C, its volume $V_C(B) \geq 0$.

Sklar's Theorem. Let G be an n-dimensional distribution function with margins $F_1, F_2, \ldots, F_n$. Then there exist an n-dimensional copula C such that, for $\mathbf{x} \in \mathbb{R}^n$

we have

$$G(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)). \tag{2.22}$$

Moreover, if $F_1, F_2, \ldots, F_n$ are continuous, then C is unique.

Definition 2. Let (X_1, X_2) be a bivariate vector of continuous random variables with marginal distribution functions F_1 and F_2. The coefficients of upper λ_U and lower λ_L tail dependence, provided that the limit $\lambda_U \in [0, 1]$ (in case of upper tail dependence) and $\lambda_L \in [0, 1]$ (in case of lower tail dependence) exist, are respectively given by the following expressions:

$$\lambda_U = \lim_{u \nearrow 1} \mathbb{P}[X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)], \tag{2.23}$$

and

$$\lambda_L = \lim_{u \searrow 0} \mathbb{P}[X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)]. \tag{2.24}$$

For a more detailed review of copula functions, refer to, e.g., Galiani (2003).

2.4.2 Connection of ASRF to Gaussian Copula

We first need to recognize that the single risk factor model in Pykhtin and Dev (2002a, 2002b) is in essence a Multivariate Gaussian Copula model. We can re-formulate the problem as follows in typical copula form:

$\tilde{T} = (T_1, T_2, \cdots, T_n)$ are default times of the portfolio of n bonds, which follow some joint distribution and each dimension has the same type of marginal distribution F (exponential, Weibull, etc.). For each bond, if $T_i < 1$ year then default happens. $P(T_i < 1) = p_i$ and LGD is μ .

To characterize the correlation structure between the default times $\tilde{T}$, we can use the Gaussian Copula

$$C(u_1, u_2, \ldots, u_n; R) = \Phi_R\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right) \tag{2.25}$$

where $R = (r_{ij})_{n\times n}$ is the covariance matrix here with $r_{ij} = 1$ and $r_{ii} = \rho$ for $i \neq j$, Φ_R is the standardized multivariate normal distribution function with covariance matrix R, and $\Phi^{-1}(u)$ denotes the inverse of standard normal cumulative distribution function.

It is obvious to see that the model above is equivalent to the ASRF model in equation (2.4). If we apply Sklar's theorem, then we get the joint default times with Gaussian Copula. Because our primary interest is in defaults and losses, not exact timing of default, so we do not need to recover default times from the copula here.

For the purpose of calculating capital required for an RMBS tranche in the context of a super-portfolio, we also need to take into account the systematic factor of the super-portfolio. According to equation (2.10),

$$Y = \sqrt{\lambda}Z + \sqrt{1-\lambda}\eta,$$

we get $corr(X_i, Z) = \sqrt{\rho\lambda}$. Thus, we can expand the previous Gaussian Copula to the following:

$$C(u_1, u_2, \ldots, u_n; \tilde{R}) = \Phi_{\tilde{R}}\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right) \tag{2.26}$$

where

$$\tilde{R} = \begin{pmatrix} R & R_{12} \\ \\ R_{21} & 1 \end{pmatrix}, \quad R_{12}^T = R_{21} = \left(\sqrt{\rho\lambda}, \cdots, \sqrt{\rho\lambda}\right)_{n\times 1}.$$

Then according to Gordy's (2003) theorems, capital for the RMBS tranche in a super-portfolio can be calculated as

$$K_q = E\left(L^{\mathrm{RMBS}} | Z = Z_{1-q}\right).$$

2.4.3 Motivation for Using an Alternative Copula

A very important property that differentiates different types of Copulas is the lower tail dependence as defined in equation (2.24). It can be shown that Gaussian Copulas have asymptotically zero tail dependence (see e.g. Galiani, 2003). Many researchers and practitioners have recognized this as a major drawback of Gaussian Copula, mainly from the perspective of valuing a structured product. For example, Malevergne and Sornette (2003) pointed out that "the Gaussian Copula hypothesis can be rejected for the dependence between pairs of commodities according to market data. Even in cases where the Gaussian Copula hypothesis is not rejected for most of the currencies and the stocks, a non-Gaussian copula, such as the Student's t Copula, cannot be rejected if it has sufficiently many 'degrees of freedom'. Then, depending on the correlation coefficient, the Student's t Copula can predict a non-negligible tail dependence which is completely missed by the Gaussian Copula assumption."

In statistical sense, valuation is focused on the center or mean of the price distribution, while capital (or VaR, more generally) which we are trying to model here is the tail metric of the loss distribution. This fact implies that Gaussian Copula may not be the most desirable when modeling the tail of a loss distribution. We would want to try other Copulas which have higher tail dependence. In banking, capital usually represents the amount of reserve that is necessary to keep a bank from insolvency in extreme conditions. In this sense, using an alternative Copula is at least an approach to be cautious or conservative.

Moreover, for a 'senior' or 'super-senior' tranche of a securitization, the degree of tail dependence will affect not only its tail loss, but also its mean or average loss. This is because of the fact that those senior tranches usually enjoy high credit support, so it is the very tail of the portfolio loss distribution that potentially could hit them. Stronger tail dependence among collateral pieces results in 'fatter' tail of the portfolio loss distribution, thus potentially making the senior tranches hit more often and more severely.

2.4.4 The Student's t Copula

In the following sections, we will use the Multivariate Student's t Copula in the place of Gaussian Copula and manifest the effect of different Copulas and tail dependence.

The Student's t Copula is defined as follows:

$$C(u_1, u_2, \ldots, u_n; R, \nu) = T_{R,\nu}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_n)\right) \tag{2.27}$$

where R is still the covariance matrix, $T_{R,\nu}$ is the standardized multivariate Student's t distribution with covariance matrix R and ν degree of freedom. $t_\nu^{-1}(u)$ denotes the inverse of Student's t cumulative distribution function.

There are three main reasons for choosing Student's t Copula for our purpose: first, the tail dependence of t Copula is not zero even if correlation is low or equal to zero. In fact, in the bivariate case we have

$$\lambda_U = 2 - 2t_{\nu+1}\left(\sqrt{\nu+1} \cdot \sqrt{\frac{1-r}{1+r}}\right)$$

where r is the correlation coefficient. As the degree of freedom increases, tail dependency decreases. $\lim_{\nu\to\infty} \lambda_U = 0$ as Student's t Copula tends to the Gaussian Copula. Secondly, despite the difference in tails, t Copula or multivariate t distribution is closely related to the multivariate normal distribution, which leads to some desirable properties such as the conditional distributions as we will see later. Last but not least, compared to other copulas with non-trivial tail dependence, such as Archimidean Copulas, the multivariate t distribution underlying the t Copula is much easier to simulate.

2.4.5 Multivariate t Distribution

According to Kotz and Nadarajah (2004), a p-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_p)^T$ is said to have the standardized p-variate t distribution with degrees of freedom ν and covariance matrix R if its joint probability density function is given by

$$f(\mathbf{x}) = \frac{\Gamma((\nu+p)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)|R|^{1/2}} \left[1 + \frac{1}{\nu}\mathbf{x}^T R^{-1}\mathbf{x}\right]^{-(\nu+p)/2}. \tag{2.28}$$

Next, we will cite properties of marginal and conditional distributions of a multivariate t distribution from Kotz and Nadarajah (2004).

Let $\mathbf{X}$ possess the p-variate t distribution with degrees of freedom ν and covariance matrix R. Consider the partitions

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

and

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

where $\mathbf{X}_1$ is $p_1 \times 1$ and R_{11} is $p_1 \times p_1$.

Then $\mathbf{X}_1$ will have the p_1-variate t distribution with degrees of freedom ν and covariance matrix R_{11}. Symmetrically, $\mathbf{X}_2$ will have the $(p-p_1)$-variate t distribution with degrees of freedom ν and covariance matrix R_{22}.

Furthermore, the conditional pdf of $\mathbf{X}_2$ given $\mathbf{X}_1$ is given by

$$\begin{aligned} f(\mathbf{x}_2|\mathbf{x}_1) &= \frac{\Gamma((\nu+p)/2)}{(\pi\nu)^{p_1/2}\Gamma((\nu+p_1)/2)}\frac{|R_{11}|^{1/2}}{|R|^{1/2}} \\ &\times\frac{[1+(1/\nu)\mathbf{x}_1^T R_{11}^{-1}\mathbf{x}_1]^{(\nu+p_1)/2}}{[1+(1/\nu)\mathbf{x}^T R_{11}^{-1}\mathbf{x}]^{(\nu+p)/2}}. \end{aligned} \tag{2.29}$$

Equation (2.29) can be rewritten as

$$\begin{aligned} f(\mathbf{x}_2|\mathbf{x}_1) &= \frac{\Gamma((\nu+p)/2)}{(\pi(\nu+p_1))^{(p-p_1)/2}\Gamma((\nu+p_1)/2)|R_{22\cdot 1}|^{1/2}} \\ &\times\left[1+\frac{1}{\nu+p_1}\frac{\mathbf{x}_{2\cdot 1}^T R_{22\cdot 1}^{-1}\mathbf{x}_{2\cdot 1}}{1+(1/\nu)\mathbf{x}_1^T R_{11}^{-1}\mathbf{x}_1}\right]^{-(\nu+p)/2} \\ &\times\left[\frac{(\nu+p_1)/\nu}{1+(1/\nu)\mathbf{x}_1^T R_{11}^{-1}\mathbf{x}_1}\right]^{(p-p_1)/2}, \end{aligned} \tag{2.30}$$

where

$$\mathbf{x}_{2\cdot 1} = \mathbf{x}_2 - R_{21}R_{11}^{-1}\mathbf{x}_1$$

and

$$R_{22\cdot 1} = R_{22} - R_{21}R_{11}^{-1}R_{12}.$$

This expression suggests that

$$\mathbf{Y}_1 = \mathbf{X}_1$$

and

$$\mathbf{Y}_2 = \frac{1}{\sqrt{diag(R_{22\cdot 1})}}\cdot\sqrt{\frac{\nu+p_1}{\nu}}\left(1+\frac{1}{\nu}\mathbf{X}_1^T R_{11}^{-1}\mathbf{X}_1\right)^{-1/2}\left(\mathbf{X}_2 - R_{21}R_{11}^{-1}\mathbf{X}_1\right) \tag{2.31}$$

are independent, that $\mathbf{Y}_1$ has the p_1-variate t distribution with degrees of freedom ν and covariance matrix R_{11}, and that $\mathbf{Y}_2$ has the $(p-p_1)$-variate t distribution with degrees of freedom $\nu+p_1$ and covariance matrix $R_{22\cdot 1}/diag(R_{22\cdot 1})$.

Therefore, although the conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1$ may not be straightforward. Equation (2.31) can be used to generate the conditional distribution via a linear transformation of a Student's t random variable.

2.4.6 Securitization Model with Student's t Copula

Now we will use the Student's t Copula described in the previous sections to model the credit risk in securitizations in the place of the original Gaussian Copula.

Model setup remains the same as in Section (2.4.2), except we will replace Gaussian copula as in Equation (2.26) with the t Copula:

$$C(u_1, u_2, \ldots, u_n; \tilde{R}, \nu) = T_{\tilde{R},\nu}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_n)\right) \tag{2.32}$$

where

$$\tilde{R} = \begin{pmatrix} R & R_{12} \\ R_{21} & 1 \end{pmatrix}, \quad R_{12}^T = R_{21} = \left(\sqrt{\rho\lambda}, \cdots, \sqrt{\rho\lambda}\right)_{n\times 1}.$$

Note: the same covariance matrix is used as in Gaussian Copula because of the fact that from a calibration point of view, the same Kendall's τ rank correlation will imply the same covariance matrix for Gaussian and t Copulas (refer to e.g. Fang & Fang, 2002).

The last dimension of the t Copula corresponds to the systematic risk factor of the super-portfolio. Due to the fact that the marginal distribution of a multivariate t distribution is still t distribution, this systematic factor Z will be a univariate t random variable.

Because of the change of correlation structure, we do not have a closed-form solution for loss of the portfolio or the expected loss of a tranche $E(L^{RMBS})$ in this case. However, it is still straightforward to simulate the defaults using Student's t Copula and use the setup in Section (2.4.2) to achieve default probability p_i, thus estimate the expected loss of each tranche.

When we consider the super-portfolio, we still have a single systematic risk factor. It is relatively simple to verify that the t distributed risk factor Z satisfies ($\mathfrak{A}$-3) and ($\mathfrak{A}$-4) in Gordy (2003) because $E[U_i|Z]$ is nondecreasing in Z for all i or each piece of collateral. Consequently, according to *Theorem* 3 of Gordy (2003), we can use

$$K_q = E\left(L^{RMBS}|Z = Z_{1-q}\right) \tag{2.33}$$

to estimate the tranche loss at the $(q \cdot 100)\%$ confidence level.

Hence, we now need to focus on the portfolio and tranche loss distribution given systematic risk factor Z. As pointed out in Section (2.4.5), we can utilize equation (2.31) to generate the conditional distribution of default times given Z,

$$\mathbf{X} = \sqrt{diag(R_{22\cdot 1})}\sqrt{\frac{\nu}{\nu + p_1}}\left(1 + \frac{1}{\nu}Z^T R_{11}^{-1} Z\right)^{1/2}\mathbf{T} + R_{21}R_{11}^{-1}Z, \tag{2.34}$$

where $\mathbf{T}$ has the n-variate t distribution with degrees of freedom $\nu+1$ and covariance matrix $R_{22\cdot 1}/diag(R_{22\cdot 1})$.

On the other hand, thanks to equation (2.33), we have transformed the estimation of LCL which is tail metric to conditional expected loss. As in most Monte

Table 2.6: Student's t Copula with different degrees of freedom

Expected Loss					
Attach	Detach	∞	100	20	10
0	3	28.7542	28.2826	27.2098	25.7022
3	6	1.1641	1.4304	2.4173	3.4178
6	10	0.0595	0.0894	0.2874	0.6085
10	100	0.0001	0.0002	0.0010	0.0040

Loss at the 99.9% Confidence Level					
Attach	Detach	∞	100	20	10
0	3	96.2201	96.8266	98.9601	99.8179
3	6	51.2746	57.3931	77.6260	91.8338
6	10	8.1151	11.8287	29.7266	54.4169
10	100	0.0132	0.0296	0.1921	0.6928

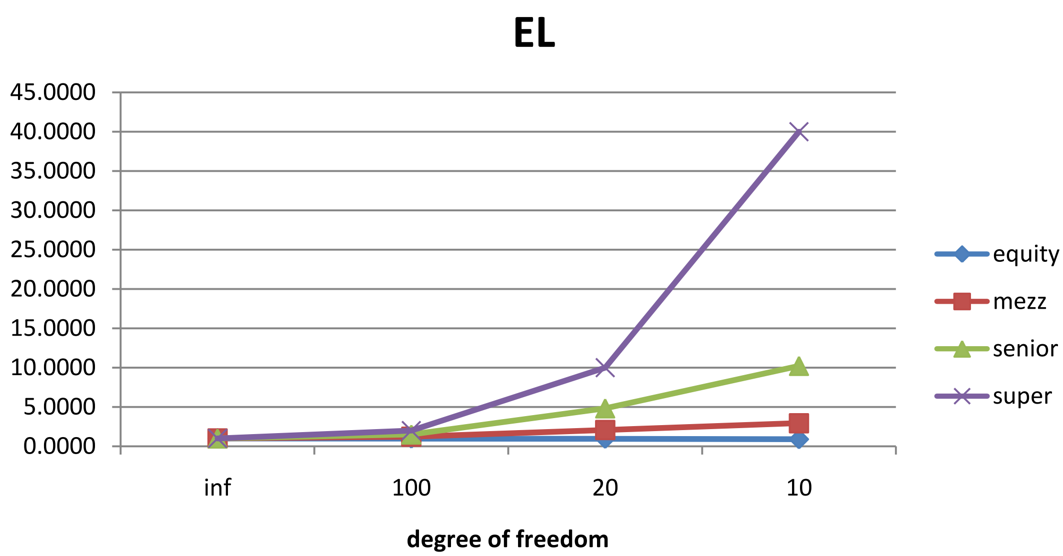
Carlo simulations, it is much easier to get a stable mean estimate than a tail estimate.

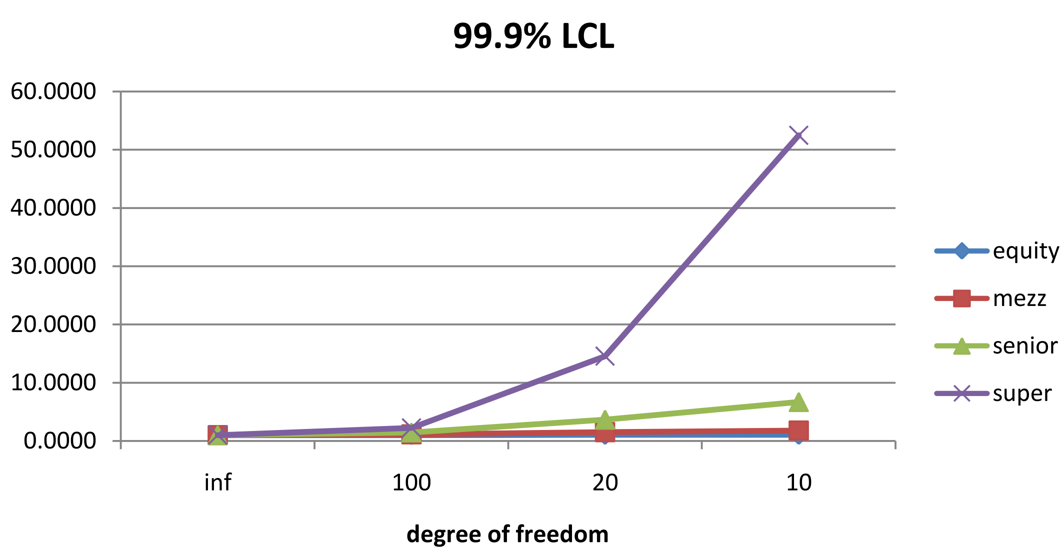
Now we have a complete Monte Carlo strategy to simulate the EL and LCL of a securitization tranche.

2.4.7 Numerical Results

We present here a simplified deal with 4 tranches: 0-3%, 3-6%, 6-10% and 10-100%. Expected loss and loss at the 99.9% confidence level are exhibited under four different Student's t Copula models: Student's t with infinite degrees of freedom, which degenerates to the Gaussian Copula, then degree of freedom at 100, 20 and 10, respectively.

Figure 2.4: EL and LCL under different degrees of freedom





It could be observed from the results in Table 2.6 that both the EL and LCL of all the tranches increase as the degree of freedom of the Student's t Copula decreases, which comes from the fact that tail dependence of Student's t Copula increases as degree of freedom decreases.

It can also be concluded that the senior tranches of the structure are affected more than junior tranches by the tail dependence. For the equity and mezzanine tranches, since the EL and LCL were already substantial even under Gaussian assumptions, the switch to Student's t Copula only increase the losses by, in some cases, just several percentage points. On the contrary, those senior tranches that have ignorable EL and LCL numbers under Gaussian Copula have sometimes startling escalations when calculated under Student's t Copula. For example, if we compare the Student's t case with 10 degrees of freedom to the Gaussian case, the super-senior tranche's (10-100%) EL hikes up about 40 times, while the LCL heightens to more than 50 times. These phenomena can also be seen from Figure 2.4.

We can appreciate the big difference that the degree of freedom makes through the Moody's ratings implied by the tranche EL's as well. The supporting senior tranche here (6-10%) has implied ratings under the four models Baa1, Baa1, Baa3 and Ba1, respectively. The super-senior tranche has Aaa, Aaa, Aa2 and A1 as its implied ratings. The drop-off of the ratings as degree of freedom decreases is more than apparent.

Chapter 3

Changepoint Approach to Survival Data

3.1 Introduction

3.1.1 Survival Analysis and Survival Data

Survival analysis is a branch of statistics that involves the modelling of time to event data; in this context, death or failure is considered an "event" in the survival analysis literature. Survival analysis attempts to answer questions such as: what is the fraction of a population which will survive past a certain time? Of those that survive, at what rate will they die or fail? Can multiple causes of death or failure be taken into account? How do particular circumstances or characteristics increase or decrease the odds of survival?

The primary objective of survival analysis is usually the "survival function" that is conventionally denoted by 'S'. The survival function is defined as

$$S(t) = \Pr(T > t)$$

where t denotes fixed time, T is a random variable denoting the time of event, and "Pr" stands for probability. That is, the survival function is the probability that the time of death is later than some specified time. The survival function is also called the 'survivor function' or 'survivorship function' in problems of biological survival.

The survival function must be non-increasing: $S(u) \leq S(t)$ if $u \leq t$. This property follows directly from $F(t) = 1 - S(t)$ being the integral of a non-negative function, as will be seen below. This reflects the notion that survival at a later age is only possible if surviving all younger ages. Given this property, the lifetime distribution function and event density, which will be defined below as F and f, are well-defined.

Related quantities are defined in terms of the survival function. The "lifetime distribution function", conventionally denoted by F, is defined as the complement of the survival function,

$$F(t) = \Pr(T \leq t) = 1 - S(t), \tag{3.1}$$

and the derivative of F (i.e., the density function of the lifetime distribution) is conventionally denoted by f,

$$f(t) = F'(t) = \frac{d}{dt}F(t). \tag{3.2}$$

f is sometimes called the "event density"; it is the rate of death or failure events per unit time. Conversely, the survival function is often also defined in terms of the distribution and density functions

$$S(t) = \Pr(T > t) = \int_t^\infty f(u)\,du = 1 - F(t). \tag{3.3}$$

The hazard function, conventionally denoted by $\lambda(t)$, is defined as the event rate at time t conditional on survival until time t or later,

$$\lambda(t)dt = \Pr(t \leq T < t + dt \,|\, T \geq t) = \frac{f(t)\,dt}{S(t)} = -\frac{S'(t)\,dt}{S(t)}. \tag{3.4}$$

The hazard function is often the focus of modeling in survival analysis. The hazard function must be non-negative, $\lambda(t) \geq 0$, and its integral over $[0, \infty)$ must be infinite, but is not otherwise constrained; the hazard function may be increasing or decreasing, nonmonotonic, or discontinuous.

The hazard function can alternatively be represented in terms of the "cumulative hazard function", conventionally denoted by $\Lambda(t)$:

$$\Lambda(t) = -\log S(t)$$

so

$$\frac{d}{dt}\Lambda(t) = -\frac{S'(t)}{S(t)} = \lambda(t) \tag{3.5}$$

$\Lambda(t)$ is called the cumulative hazard function because the preceding definitions together imply

$$\Lambda(t) = \int_0^t \lambda(u)\,du$$

which is the "accumulation" of the hazard over time.

Censoring is a form of missing data problem which is common in survival analysis. Ideally, both the birth and death dates of a subject are known, in which case the lifetime is known. If it is known only that the date of death is after some date, this is called right censoring. Right censoring will occur for those subjects whose birth date is known but who are still alive when they are lost to follow-up or when the study ends. A lot of the methodologies in survival analysis deals with the problem of censoring. We deal primarily with right censoring in this thesis.

With the presence of censoring, we introduce the notion of censored data. Let T_i and C_i be the survival time and censoring time for ith individual, respectively. The observations are the observed times $Y_i = T_i \wedge C_i$ and censoring indicators $\delta_i = 1_{(T_i \leq C_i)}$.

The classical method to estimate the survival function of the failure time non-parametrically is the Kaplan-Meier estimator. If we sort the observed failure times as $t_1 < t_2 \cdot < t_k$ and let d_i and n_i denote the number of failures at time t_i and the number of individuals at risk at a time just prior to t_i, then the Kaplan-Meier estimator of the survival function is

$$\hat{F}(t) = \prod_{i|t_i \leq t} \frac{n_i - d_i}{n_i}$$

This was first derived by Kaplan and Meier (1958) and also referred to as the product limit estimate of the survivor function.

A related method to estimate the cumulative hazard function non-parametrically is the Nelson-Aalen estimator. The estimator is given as follows

$$\hat{\Lambda}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$$

3.1.2 Cox Model and Partial Likelihood

The present thesis focuses on a parametric regression problem instead. In this regression setting, in addition to Y_i and δ_i, we also observe covariates X_i. By regression modeling, we mean to relate failure time T_i to a covariate specific hazard function. In this connection, the most commonly used statistical model is the well-known Cox proportional hazards model (Cox, 1972). If X_i represents treatment assignment, the Cox model specifies that the treatment has a multiplicative effect on the subject's hazard rate. For example, a drug may halve one's immediate probability of stroke. This is in contrast to additive hazards models, wherein a treatment may increase one's hazard by a fixed amount which is independent of baseline.

We can consider the Cox model as consisting of two parts: the baseline hazard function, often denoted by $\Lambda_0(t)$, describing how hazard (risk) changes over time at 'baseline' levels of covariates; and the effect parameters, describing how the hazard varies in response to explanatory covariates. A typical medical example would include as covariates, treatment assignment as well as patient characteristics to reduce variability and/or control for confounding.

Let $\lambda(t|x)$ be the hazard function when the covariate $X = x$, which can be multi-dimensional. The Cox model takes the form

$$\lambda(t|x) = e^{\beta x}\lambda_0(t), \tag{3.6}$$

where $\lambda_0(t)$ is the baseline hazard rate. Treatment is a dichotomous variable included in x and its corresponding coefficient β is called the 'treatment effect'.

Suppose there are n independent subjects, each associated with a $p \times 1$ vector of covariates $X_i = (x_{i1}, \ldots, x_{ip})'$ $(1 \leq i \leq n)$. Let T_i and C_i be the survival time and censoring time of the ith subject respectively. We assume that $\{(T_i, C_i, X_i),\ 1 \leq i \leq n\}$ are independent and identically distributed; also that the $\{T_i\}$ and $\{C_i\}$ are conditionally independent of each other, given the $\{X_i\}$. We observe pairs $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$, where $Y_i = T_i \wedge C_i$ and $\delta_i = 1_{(T_i \leq C_i)}$. So the data consist of n triplets $\{(Y_i, \delta_i, X_i),\ 1 \leq i \leq n\}$.

Sir David Cox observed that if the proportional hazards assumption holds (or, is assumed to hold) then it is possible to estimate the effect parameter(s) without any consideration of the baseline hazard function. In other words, the Cox model can be analyzed using this approach called "partial likelihood",

$$L(\beta) = \prod_{j=1}^{k} \frac{\exp[X_j'\beta]}{\sum_{l \in \mathcal{R}(t_j)} \exp[X_l'\beta]}.$$

Here $\mathcal{R}(t)$ is the set of individuals at risk of failure at time $t-$. An estimator of the treatment effect $\hat{\beta}$ can be obtained by maximizing the partial likelihood function $L(\beta)$.

Following the estimated $\hat{\beta}$, the baseline hazard function can also be estimated by applying the Nelson-Aalen estimator $\hat{\Lambda}(t) = \int_0^t d\hat{\Lambda}_0(u)$, where $d\hat{\Lambda}_0(u)$ is 0 except at the observed failure times t_i, where it takes the value

$$d\hat{\Lambda}_0(u) = d_i \left\{ \sum_{l \in \mathcal{R}(t_i)} \exp[X_l' \hat{\beta}] \right\}^{-1}.$$

In the simplest case of stationary coefficients, for example, a treatment with a drug may, say, halve a subject's hazard at any given time t, while the baseline hazard may vary. Note however, that the covariate is not restricted to binary predictors. In the case of a continuous covariate x, the hazard responds logarithmically; each unit increase in x results in proportional scaling of the hazard. Typically under the fully-general Cox model, the baseline hazard is "integrated out", or heuristically removed from consideration, and the remaining partial likelihood is maximized. The effect of covariates estimated by any proportional hazards model can thus be reported as hazard ratios.

3.1.3 Lag Effect and Changepoint

In survival analysis, it is often not unreasonable to assume that a treatment effect does not affect the risk of failure until after a certain time lag, which is quite common in clinical trials for disease prevention or treatment. Zucker and Lakatos (1990) described several such situations. There is also the possibility of a "saturation effect", where the treatment stops to affect the risk of failure after a certain period of time.

We will consider a simple way to characterize the lag effect. It means that the treatment effect stays at zero until a thresholding time, at which it takes on a nonzero value. This means that the hazard rate takes a jump (or dive) at that time. So the statistical problem of interest would be to test whether such a threshold or 'changepoint' exists, and to estimate the location of the changepoint and the treatment effect before and after the changepoint. The treatment effect is initially zero by assumption in this case. Such knowledge would be of great value in relevant clinical trials.

Mathematically, it is very natural to extend the problem to the 'saturation effect' where the treatment effect changes from nonzero to zero at a certain threshold, or situations where the treatment effect is not necessarily zero at the beginning, but just takes a jump of a certain magnitude at the changepoint. We can also generalize to multiple changepoints so that such jump could take place multiple times throughout the whole time period of interest.

In statistical literatures, in terms of hypothesis testing, Zucker and Lakatos (1990) studied two-sample weighted logrank statistics for comparing survival curves that are robust against a class of unknown treatment lag effects. Liang, Self, and Liu (1990) presented a modification of the Cox (1972) proportional hazards model to include a changepoint. They proposed a test of the null hypothesis that there is no changepoint based on the maximal score statistic. Luo, Turnbull, and Clark (1997) used the same model and proposed a partial likelihood ratio test for testing the hypothesis that the changepoint is equal to a given value.

For estimation of lag effects, Luo et al. (1994) and Luo (1996) used the maximum partial likelihood estimators and obtained the asymptotic distributions for the estimators of the changepoint and regression coefficients.

Luo's model (1994, 1996, 1997) is a simple extension of the Cox proportional hazards model. Their testing and estimation methodologies are for the most part standard likelihood-based methods such as the likelihood ratio test and the maximum (partial) likelihood estimator. Unlike in the standard case, the test statistic they arrived at is not of a normal distribution. Calculating the p-value would involve simulation of random-walks. See Luo, Turnbull, and Clark (1997) for details. Their maximum likelihood estimator of the changepoint does not have a standard asymptotic distribution either. See Luo (1996) for more discussion.

In this part of the thesis, we consider a model similar to the one used by Liang et al. (1990) and Luo et al. (1994), with the multiple changepoint detection technique developed by Lai and Xing (2011). The next section gives a brief introduction to the modeling techniques of changepoint problems.

3.1.4 Changepoint Modeling

In general, the changepoint problems in statistics are based on a series of independent observations from the family of distributions f_{θ_i} where the θ_i are unknown parameters that are piecewise constant. The problems of interest are the testing and estimation of the parameter values and the changepoints (where the parameter changes value).

There is an extensive literature on the case where θ_i can undergo at most one change, both from the frequentist approach (e.g., Hinkley, 1970) and the Bayesian approach (e.g., Carlin, Gelfand, & Smith, 1992). There has been difficulty extending the methodology to the multiple changepoint setting because of the computational complexity.

From the frequentist perspective, Bai (1997a, 1997b), Bai and Perron (1998, 2003), and Qu and Perron (2007) considered regression models with multiple changepoints, using dynamic programming to compute the least square estimates of the piecewise constant regression parameters when it is assumed that there are $k(\geq 2)$ changepoints. An alternative approach that is computationally more convenient especially when k is not small is the binary segmentation procedure proposed by Vostrikova (1981) and refined by Olshen et al. (2004). The choice of k for this approach can be carried out by a model selection criterion by applying Schwartz's Bayesian Information Criteria (BIC) as in Yao (1988). Siegmund (2004) and Zhang and Siegmund (2006) pointed out that the likelihood functions do not satisfy the regularity conditions of the BIC and proposed modifications of the BIC in changepoint problems.

The Bayesian approach to multiple changeponts dates back to the work of Chernoff and Zacks (1964). McCulloch and Tsay (1993) extended Chernoff-Zacks model and used the Gibbs sampler to approximate the posterior distribution of the time-varying parameters. Barry and Hartigan (1992, 1993) proposed a product partition model as the prior distribution for the sequence of the piecewise constant parameters and used the Gibbs sampler to estimate the posterior means of the parameters. Subsequent developments of the Bayesian approach typically utilized the reversible jump Markov Chain Monte Carlo (MCMC) method by Green (1995), or Gibbs sampling used in conjunction with Metropolis-Hastings steps, such as Liu and Lawrence (1999), Wang and Zivot (2000) and Chib, Nardari and Shephard (2002). All these methods assume conjugate priors for the parameters and use simulation-based inference via MCMC algorithms.

Lai and Xing (2011) considers a Bayesian model for multiple changepoints in a multiparameter exponential family. The model has attractive statistical and computational properties. It produces explicit recursive formulas for the Bayesian estimates of the piecewise constant parameters instead of depending on simulation techniques. Although the approach assumes a parametric model and superimposes on it a Bayesian changepoint model, the assumed model is only used as a *working model* to derive the Bayesian smoothers and the frequentist segmentation procedures. Their model is able to tackle both Bayesian and frequentist problems efficiently provides a general methodology for multiple changepoint problems.

In this part, we propose an approach to modeling changepoint problems in Cox model that is parallel to Lai and Xing (2011) and the procedure shares many desirable properties. However, Lai and Xing's model is built under a time series framework while we are dealing with survival data in this part. There are two major differences. Firstly, all observations are made within finite time in our setting. Secondly, we will have only one observation (failure time) for each individual but multiple individuals. Useful asymptotic properties can be established as the number of individuals increases.

3.2 Model Specification

3.2.1 Changepoint Framework

3.2.1.1 Independent and Identically Distributed Case

Following Lai and Xing (2011), we consider a multiparameter exponential family of densities

$$f_\theta(y) = \exp\{\theta' y - \psi(\theta)\} \tag{3.7}$$

with respect to some measure ν on $\mathbb{R}^d$, and the prior density of π. $\{y_t\}$ $(t = 1, \ldots, n)$ are independent and indentically distributed (i.i.d.) observations from the densities.

Suppose that, instead of being time-invariant, the parameter vector θ_t may undergo occasional changes such that for $t > 1$, the indicator variables

$$I_t := \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \tag{3.8}$$

are independent Bernoulli random variables with $P(I_t = 1) = p$, where p stands for the probability of change. If $p = 0$, the model reverts back to the time-invariant case. Again, this serves only as a working assumption for our derivation under the Bayesian setup. The procedures and results are not reliant on the assumption.

When there is a parameter change at time t (i.e., $I_t = 1$), the changed parameter θ_t is assumed to be sampled from the same density π. In other words, θ_t is a stepwise constant function with respect to time t and its value jumps at the changepoints. Consequently, y_t will not be identically distributed any more.

3.2.1.2 Time Series Case

To illustrate in a time series setup, we take the simple AR(1) as an example. We assume y_t follows an AR(1) process with no intercept term

$$y_t = \theta \cdot y_{t-1} + \epsilon_t,$$

where $t > 1$ and ϵ_t is a white noise process with density ϕ.

In a conventional time series, θ is assumed to be time-invariant. Given it is an AR(1) process, the likelihood of the time series $y_1, \ldots, y_n$ can be written as

$$\begin{aligned} f(y_1, \ldots, y_n) &= f(y_n|y_{n-1}) \cdots f(y_2|y_1) f(y_1) \\ &= \phi(y_n - \theta y_{n-1}) \cdots \phi(y_2 - \theta y_1) f(y_1). \end{aligned} \tag{3.9}$$

Here, we again assume that θ_t may undergo occasional changes such that for $t > 1$, the indicator variables

$$I_t := \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \tag{3.10}$$

are independent Bernoulli random variables with $P(I_t = 1) = p$.

The AR(1) process becomes $y_t = \theta_{t-1} \cdot y_{t-1} + \epsilon_t$. Consequently, the likelihood expression becomes

$$\begin{aligned} f(y_1, \ldots, y_n) &= f(y_n|y_{n-1}) \cdots f(y_2|y_1) f(y_1) \\ &= \phi(y_n - \theta_{t-1} y_{n-1}) \cdots \phi(y_2 - \theta_1 y_1) f(y_1). \end{aligned} \tag{3.11}$$

3.2.1.3 Changepoint Model on Exponential Family and Conjugate Prior

Consider a multiparameter exponential family of densities

$$f_\theta(y) = \exp\{\theta' y - \psi(\theta)\} \tag{3.12}$$

with respect to some measure ν on $\mathbb{R}^d$, and the prior density of π on $\Theta := \{\theta : \int e^{\theta' y} d\nu(y) < \infty\}$ given by

$$\pi(\theta; a_0, \mu_0) = c(a_0, \mu_0) \exp\{a_0 \mu_0' \theta - a_0 \psi(\theta)\}, \quad \theta \in \Theta, \tag{3.13}$$

where $1/c(a_0, \mu_0) = \int_\Theta exp\{a_0 \mu_0' \theta - a_0 \psi(\theta)\} d\theta$ and $\mu_0 \in (\nabla \psi)(\Theta)$, in which ∇ denotes the gradient vector of partial derivatives. The posterior density of θ given by the observations $y_1, \ldots, y_m$ drawn from f_θ is

$$\pi\big(\theta; a_0 + m, \frac{a_0 \mu_0 + \sum_{i=1}^m y_i}{a_0 + m}\big); \tag{3.14}$$

see Diaconis and Ylvisaker (1979). Thus, (3.13) is a conjugate family of priors and

$$\int_\Theta f_\theta(y) \pi(\theta; a, \mu) d\theta = \frac{c(a, \mu)}{c(a+1, (a\mu + y)/(a+1))}. \tag{3.15}$$

As before, we assume that the parameter vector θ_t may undergo occasional changes such that for $t > 1$, the indicator variables

$$I_t := \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \tag{3.16}$$

are independent Bernoulli random variables with $P(I_t = 1) = p$. When there is a parameter change at time t (i.e., $I_t = 1$), the changed parameter θ_t is assumed to be sampled from π.

The simplicity of the conjugate priors allows for explicit formulas for the sequential (filtering) estimates $E(\mu_t|Y_{1,t})$ and for the fixed-sample (smoothing) estimates $E(\mu_t|Y_{1,n})$, where $\mu_t = \nabla\psi(\theta_t)$ and $Y_{i,j}$ denotes $(y_i, \ldots, y_j)$ for $i \leq j$.

An important ingredient in the development of these explicit formulas is the most recent changepoint R_t up to time t, i.e., $R_t = \max\{s \leq t : I_s = 1\}$. In other words, R_t is the last changepoint on or before t, implying that there are no other changepoints between K_t and t and all observations between the two points share the same parameter θ.

Let $p_{it} = P(K_t = i|Y_t)$. Using Bayes' formula, note that

$$f(\theta_t|Y_{1,t}) = \sum_{i=1}^{t} p_{it} f(\theta_t|Y_{i,t}, K_t = i). \tag{3.17}$$

Here $Y_{1,t}$ can be replaced by $Y_{i,t}$ on the right-hand side because observations before the most recent changepoint do not carry information about the current parameter θ_t.

Given $R_t = i$, $Y_{i,t} = (y_i, \dots, y_t)$ are consecutive observations sharing the same parameter θ_t. It follows from (3.14) that

$$f(\theta_t | Y_{i,t}, R_t = i) = \pi(\theta_t; a_0 + t - i + 1, \bar{Y}_{i,t}), \tag{3.18}$$

where $\bar{Y}_{i,t} = (a_0\mu_0 + \sum_{k=i}^{j} y_k)/(a_0 + j - i + 1)$ for $j \geq i$. Combining the two equations above yields

$$f(\theta_t | Y_{1,t}) = \sum_{i=1}^{t} p_{it} \pi(\theta_t; a_0 + t - i + 1, \bar{Y}_{i,t}). \tag{3.19}$$

A recursive formula for p_{it} can be obtained by noting that $\sum_{i=1}^{t} p_{it} = 1$ and

$$p_{it} \propto p_{it}^* := \begin{cases} pf(y_t | I_t = 1) & \text{if } i = t, \\ \\ (1-p)p_{i,t-1} f(y_t | Y_{i,t-1}, K_t = i) & \text{if } i \leq t - 1. \end{cases} \tag{3.20}$$

Plugging in $f(y_t | Y_{i,t-1}, K_t = i) = \int f_{\theta_t}(y_t) f(\theta_t | Y_{i,t-1}, K_t = i) d\theta_t$ together with (3.15) and (3.18), we get

$$p_{it}^* := \begin{cases} \dfrac{p\pi_{0,0}}{\pi_{t,t}} & \text{if } i = t, \\ \\ (1-p)p_{i,t-1} \dfrac{\pi_{i,t-1}}{\pi_{i,t}} & \text{if } i \leq t - 1, \end{cases} \tag{3.21}$$

where $\pi_{0,0} := c(a_0, \mu_0)$ and $\pi_{i,j} := c(a_0 + j - i + 1, \bar{Y}_{i,j})$.

To derive $E(\mu_t | Y_{1,n})$, we utilize Bayes Theorem to combine the forward filter $\theta_t | Y_t$ and the backward filter $\theta_t | Y_{t+1,n}$. The backward filter is obtained by reversing time, noting that the $\widetilde{I}_t = 1_{\{\theta_t \neq \theta_{t+1}\}}$ are still independent Bernoulli. We define the very next changepoint as the time-reversed counterpart of R_t, $\widetilde{R}_t := \min\{s > t : \widetilde{I}_s = 1\}$ and $q_{t+1,j} = P(\widetilde{R}_t = j | Y_t)$.

The backward (time-reversed) filter can be expressed as

$$f(\theta_t|Y_{t+1,n}) = p\pi(\theta_t; a_0, \mu_0) + (1-p)\sum_{j=t+1}^{n} q_{j,t+1}\pi(\theta_t; a_0 + j - t, \bar{Y}_{t+1,j}), \tag{3.22}$$

where $q_{jt} \propto q_{jt}^*$, $\sum_{j=t}^{n} q_{jt} = 1$ and

$$q_{jt}^* := \begin{cases} \dfrac{p\pi_{0,0}}{\pi_{t,t}} & \text{if } j = t, \\ (1-p)q_{j,t+1}\dfrac{\pi_{t+1,j}}{\pi_{tj}} & \text{if } j > t. \end{cases} \tag{3.23}$$

By Bayes Theorem,

$$f(\theta_t|Y_{1,n}) \propto \frac{f(\theta_t|Y_{1,t})f(\theta_t|Y_{t+1,n})}{\pi(\theta; a_0, \mu_0)}. \tag{3.24}$$

Combining the backward filter (3.22) with the forward filter (3.19), and noting that

$$\pi(\theta; a_0 + t - i + 1, \bar{Y}_{i,t})\frac{\pi(\theta; a_0 + j - t, \bar{Y}_{t+1,j})}{\pi(\theta; a_0, \mu_0)} = \frac{\pi_{it}\pi_{t+1,j}}{\pi_{ij}\pi_{00}}\pi(\theta; a_0 + j - i + 1, \bar{Y}_{i,j}),$$

we obtain from (3.24) that

$$f(\theta_t|Y_{1,n}) = \sum_{1\le i\le t\le j\le n} \beta_{ijt}\pi(\theta_t; a_0 + j - i + 1, \bar{Y}_{i,j}), \tag{3.25}$$

where $\beta_{ijt} = \beta_{ijt}^*/P_t$, $P_t = p + \sum_{1\le i\le t<j\le n} \beta_{ijt}^*$, and

$$\beta_{ijt}^* := \begin{cases} pp_{it} & \text{if } i \le t = j, \\ (1-p)p_{it}q_{j,t+1}\dfrac{\pi_{it}\pi_{t+1,j}}{\pi_{ij}\pi_{00}} & \text{if } i \le t < j. \end{cases} \tag{3.26}$$

It follows from (3.26) that

$$P(I_{t+1} = 1|Y_{1,n}) = \frac{p}{P_t}, \qquad E(\mu_t|Y_{1,n}) = \sum_{1\le i\le t\le j\le n} \beta_{ijt}\bar{Y}_{i,j}. \tag{3.27}$$

3.2.2 Cox Model with Changepoints

The changepoint problem in survival analysis, specifically in Cox model shares many similarities with the general changepoint problem. However, a major difference conceptually is the fact that the changepoint is not modeled on the response variable "time" directly as in most other models. Instead, changepoint is modeled on the hazard function, which is an instantaneous rate of the event (or failure) happening. Thus, the changepoint affects time indirectly through the hazard function. In other words, we will make the assumption that the coefficient β in Cox model may undergo occasional changes and is a piecewise constant function with respect to time.

Another important difference comes from the fact that Cox model is a semi-parametric models. Instead of the full likelihood, we have to rely on the partial likelihood for inference of the coefficient. In order to be analogous to the general method, we first introduce a segmentation of the time axis $0 < t_1 < \ldots < t_j < \ldots < t_k < \ldots < t_K < \infty$. We assume that the coefficient stays constant within each time segment $[t_j, t_{j+1}],\ 1 \leq j < K$ so that change could only occur on a finite number of time points.

Under the standard setup of Cox model, suppose there are n independent subjects, each associated with a $p \times 1$ vector of covariates $X_i = (x_{i1}, \ldots, x_{ip})'$ $(1 \leq i \leq n)$. Let T_i and C_i be the survival time and censoring time of the ith subject, respectively. We assume that $\{(T_i, C_i, X_i),\ 1 \le i \le n\}$ are independent and identically distributed; also that the $\{T_i\}$ and $\{C_i\}$ are conditionally independent of each other, given the $\{X_i\}$. We observe pairs $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$, where $Y_i = T_i \wedge C_i$ and $\delta_i = 1_{(T_i \le C_i)}$. So the data consist of n triplets $\{(Y_i, \delta_i, X_i),\ 1 \le i \le n\}$.

With respect to the time segmentation, we here define Y_{jk} as the set of failure time observations made between t_j and t_k:

$$Y_{jk} = \{\ Y_i \mid Y_i \in [t_j, t_k],\ \delta_i = 1\ \} \tag{3.28}$$

Note that in general, there could be multiple failure times within each segment, or there could be zero failure time with each segment.

As a result, for time segment $[t_j, t_k]$, we have partial likelihood within the segment $f(Y_{jk}|\beta)$ defined as follows:

$$f(Y_{jk}|\beta) = \prod_{i:\, Y_i \in Y_{jk}} \frac{\lambda_{x_i}(Y_i)}{\sum_{i' \in \mathcal{R}_i} \lambda_{x_{i'}}(Y_{i'})} = \prod_{i:\, Y_i \in Y_{jk}} \frac{e^{\beta X_i}}{\sum_{i' \in \mathcal{R}_i} e^{\beta X_{i'}}} \tag{3.29}$$

As the usual notation in Cox models, $\mathcal{R}_i = \{i' | Y_{i'} \ge Y_i\}$ denotes the risk set at time Y_i, meaning all the subjects at risk at failure time Y_i; $\lambda_x(t) = e^{\beta x}\lambda_0(t)$ is the hazard function for a subject with covariate $X = x$. Here $\lambda_0(t)$ is an arbitrary unknown baseline hazard function. Intuitively, $f(Y_{jk}|\beta)$ is the product of the partial likelihood terms for every failure time between t_j and t_k.

As discussed before, to accommodate for the existence of changepoints, we now make the assumption that β is not time-invariant and may undergo occasional changes. The proportional hazards model with time-varying coefficient has the form

$$\lambda(t|x) = e^{\beta(t)x}\lambda_0(t). \tag{3.30}$$

$\beta(t)$ denotes the regression coefficient that is time dependent. (Following the Bayesian setup in Lai et al. (2005), we assume that $\beta(t)$ forms a variant of compound poisson process with rate r:

$$\beta(t) := D_{N(t)} \tag{3.31}$$

$\{N(t), t \geq 0\}$ is a Poisson process with rate r, and $\{D_i, i \geq 0\}$ are i.i.d. random variables with known distribution Π, which are also independent of $\{N(t), t \geq 0\}$.)

Please note that the Bayesian setup in this section allows a simple algorithm to be developed, but the results will not be dependent on the Bayesian setup. As shown later in the section, the results will have frequentist interpretations as well as Bayesian interpretations.

For notation simplicity, when there is no ambiguity, we will use the notations β_k and I_k as follows

$$\beta_k = \beta(t_k),\ I_k = \mathbf{1}_{\{\beta_k \neq \beta_{k-1}\}}.$$

As a result, $P(I_k = 1) = P(\beta_k \neq \beta_{k-1}) = 1 - e^{-r(t_k - t_{k-1})}$. By utilizing the most recent changepoint $R_k = \{j | j \leq k; I_j = 1\}$, we can derive the posterior distribution of β_k given Y_{1k} (forward filter):

$$f(\beta_k|Y_{1k}) = \sum_{j=1}^{k} P(R_k = j|Y_{1k}) f(\beta_k|R_k = j, Y_{1k}) = \sum_{j=1}^{k} p_{jk} f(\beta_k|R_k = j, Y_{1k}) \tag{3.32}$$

and

$$p_{jk} \propto p_{jk}^* = \begin{cases} p_k f_{kk} & \text{if } j = k, \\ \\ (1-p_k)p_{j,k-1}\dfrac{f_{jk}}{f_{j,k-1}} & \text{if } j < k, \end{cases} \tag{3.33}$$

where

$$f_{jk} = f(Y_{jk}|C_{jk}) = \int_{-\infty}^{\infty} f(Y_{jk}|\beta)\Pi(\beta)d\beta, \tag{3.34}$$

and

$$C_{jk} = \{I_j = 1, I_{j+1} = 0, \cdots I_k = 0, I_{k+1} = 1\}. \tag{3.35}$$

Similarly, by utilizing the very next changepoint $\widetilde{R}_k = \{l|l > k; \widetilde{I}_l = 1\}$ where $\widetilde{I}_k = \mathbf{1}_{\{\beta_k \neq \beta_{k+1}\}}$, we can calculate the posterior distribution of β_{k+1} given $Y_{k+1,K}$ (backward filter):

$$\begin{aligned} f(\beta_{k+1}|Y_{k+1,K}) &= \sum_{l=k+1}^{K} P(\widetilde{R}_k = l|Y_{k+1,K}) f(\beta_{k+1}|\widetilde{R}_k = l, Y_{k+1,K}) \\ &= \sum_{l=k+1}^{K} q_{k+1,l} f(\beta_{k+1}|\widetilde{R}_k = l, Y_{k+1,K}) \end{aligned} \tag{3.36}$$

and

$$q_{k+1,l} \propto q_{k+1,l}^* = \begin{cases} p_{k+1} f_{k+1,k+1} & \text{if } l = k+1 \\ \\ (1-p_{k+1})q_{k+2,l}\dfrac{f_{k+1,l}}{f_{k+2,l}} & \text{if } l > k+1. \end{cases} \tag{3.37}$$

Combining the forward and backward filters gives us the posterior distribution of each β_k given the whole set of observations Y_{1K} (smoother):

$$f(\beta_k|Y_{1K}) \propto f(\beta_k|Y_{1k})f(\beta_k|Y_{k+1,K})/\Pi(\beta_k)$$

$$= f(\beta_k|Y_{1k})[p_{k+1} + (1-p_{k+1})f(\beta_{k+1}|Y_{k+1,K})/\Pi(\beta_k)] \tag{3.38}$$

$$= \sum_{j=1}^{k} p_{k+1}p_{jk}f(\beta_k|Y_{1K}, C_{jk}) + \sum_{1\le j\le k<l\le K} (1-p_{k+1})p_{jk}q_{k+1,l}\frac{f_{jl}}{f_{jk}f_{k+1,l}}f(\beta_k|Y_{1K}, C_{jl})$$

Here $f(\beta_k|Y_{1K}, C_{jk})$ and $f(\beta_k|Y_{1K}, C_{jl})$ can be easily calculated because they are essentially posterior distributions of β_k given a portion of the data without changepoint. In the same way, the posterior changepoint probability $\widehat{p}_{k+1}$ at each failure time can be derived as the posterior mean of I_k:

$$\widehat{p}_{k+1} = \frac{p_{k+1}}{p_{k+1} + \sum_{1\le j\le k<l\le K}(1-p_{k+1})p_{jk}q_{k+1,l}f_{jl}/(f_{jk}f_{k+1,l})} \tag{3.39}$$

If one single changepoint is assumed, this model is equivalent to the lag model used by Luo et al. (1996, 1997). However, our model is more general in the sense that the number of changpoints is not pre-specified in the model. With the procedure described below, all of the existing changepoints will be detected.

As in Lai et al. (2005), by utilizing the most recent change-time, we can get the forward filter, backward filter, then the posterior distribution for I_t and $\beta(t), t = t_1, \ldots, t_K$. The key thing to note here is that in $f_{jk} = \int_{-\infty}^{\infty} f(Y_{jk}|\beta)\Pi(\beta)d\beta$, because of the nonparametric term $\lambda_0(t)$ in our model, we cannot write out the full likelihood here. So instead of the full likelihood, we can use partial likelihood for $f(Y_{kl}|\beta)$:

$$f(Y_{jk}|\beta) = \prod_{i:\, Y_i\in Y_{jk}} \frac{\lambda_{x_i}(Y_i)}{\sum_{i'\in\mathcal{R}_i}\lambda_{x_{i'}}(Y_{i'})} = \prod_{i:\, Y_i\in Y_{jk}} \frac{e^{\beta X_i}}{\sum_{i'\in\mathcal{R}_i} e^{\beta X_{i'}}} \tag{3.40}$$

The previous discussion is based on "Fixed Segmentation" where the segmentation points are pre-determined. Given all the observations Y_i, we can define the "Natural Segmentation", which is using all the failure times to segment the time axis instead of arbitrary time points. Thus, the segmentation points are $\{t_j\} = \{Y_i|\delta_i = 1\}$. Consequently, there is only one failure time in each segment. The "Natural Segmentation" is the finest segmentation possible for inference on β, for the reason that there is no information about the coefficient between two adjacent failure times.

One of the major differences between this Cox model with Changepoints and the Changepoint Model in the previous section is the fact that there is no conjugate prior for the partial likelihood in the Cox model. As a result, there is no explicit formula for the integrals f_{ij} as in (3.34) and they have to be carried out numerically in general. If a normal prior is assumed, one way to carry out the numerical integration is techniques such as Gauss-Hermite quadrature.

Under normal prior, we can also utilize quadratic approximation to avoid the need of numerical integration. If there are enough observations between i and j, we can perform a quadratic expansion of the partial likelihood function around the estimator $\hat{\beta}$ as follows:

$$f(Y_{jk}|\beta) = e^{\frac{1}{2}(\beta-\hat{\beta})' J(\hat{\beta})(\beta-\hat{\beta})}, \tag{3.41}$$

where $J(\cdot)$ is the Fisher information matrix and $\hat{\beta}$ is the maximum partial likelihood estimator. Inserting (3.41) back into (3.34) gives the marginal likelihood as integration of the product of two normal densities:

$$f_{jk} = f(Y_{jk}|C_{jk}) = \int_{-\infty}^{\infty} e^{\frac{1}{2}(\beta-\hat{\beta})' J(\hat{\beta})(\beta-\hat{\beta})} \Pi(\beta) d\beta, \tag{3.42}$$

which yields an explicit formula as the general case in the previous section.

3.2.3 Estimation Procedure

3.2.3.1 Fixed Segmentation and Natural Segmentation

If prior knowledge is available with respect to the location of changepoints, or if changepoints can only occur at a number of time points due to certain experimental designs, the "Fixed Segmentation" is recommended for changepoint detection.

For example, if it was known that a certain drug has a lag effect of 3-6 months and a saturation effect after 5-6 years, the the segmentation can be set up on fixed time points such as (3 month, 4 month, 5 month, 6 month, 5 years, 5.25 years, 5.5 years, 6 years) and the one changepoint can be identified for each effect.

It will be shown in the next chapter theoretically that the changepoints identified in this setup has nice asymptotic properties.

If no prior knowledge of the changepoints is present, the recommended method is the "Natural Segmentation". It is the finest segmentation possible with the data given, and as shown by simulation studies afterwards, it can detect changepoints effectively when the sample size is reasonably large.

3.2.3.2 Single Changepoint

If it is assumed that there is only one changepoint, the estimation procedure for the changepoint k_c is then simply finding the point with the maximum posterior changepoint probability:

$$\{k_c : \hat{p}_{k_c} = \max_{1 \leq k \leq K} (\hat{p}_k)\}.$$

As will be shown in the next chapter, the posterior changepoint probability on the true changepoint converges to 1, while the probability on other points converges to 0. As a result, this procedure will always find the true changepoint asymptotically.

3.2.3.3 Multiple Changepoints

If multiple changepoints are assumed, one has to set a cutoff threshold for posterior changepoint probability, for example, 0.5. Any point with $\hat{p}_k$ above the threshold is identified as a changepoint. With the asymptotic properties, true changepoints will be identified given enough observations.

However, in applications such as the lag effect and saturation effect in the example above, the changepoints often have direct scientific interpretations, resulting in more efficient ways of identifying the changepoints. In the lag effect and saturation effect example above, the procedure will be to identify one changepoint indentified between 3 to 6 months, another one between 5 to 6 years.

3.2.3.4 Estimation of β

There are two ways to estimate β. The Bayesian estimate would be the posterior mean as shown in (3.38), which yields a β estimate for each time point. An abrupt change in the estimate can be observed where a changepoint is most likely.

The other way to estimate $\hat{\beta}$ is from the frequentist perspective after the changepoints are identified. For example, if one changepoints is identified, the changepoints will partition the data into two sections. The maximum partial likelihood estimator of each section $\hat{\beta}_1$ and $\hat{\beta}_2$ will be the estimate coefficient for each section.

3.3 Proof of Consistency

We will prove the consistency of the changepoint estimates of the Cox model under the "Fixed Segmentation" setup in the chapter. It will be shown that as the number of observations goes to infinity, the posterior changepoint probabilities $\hat{p}_k$ on the true changepoints converge to 1, while the posterior changepoint probabilities on non-changepoints converge to 0.

We will start by looking at the single changepoint scenario and the generalize to the scenario of multiple changepoints.

3.3.1 Single Changepoint

Following the "fixed segmentation", We segment time into intervals of $[0, t_1]$, $[t_1, t_2]$,...,$[t_{m-1}, t_m]$, and assume t_{k_c+1} $(0 < k_c < m)$ is the true changepoint in terms of the parameter of interest θ. Let $\{Y_{i,1}, i = 1, 2...n\}$ denote the n observations made between $[0, t_1]$. Let $\{Y_{i,k}, i = 1, 2...n\}$ denote the n observations made between $[t_{k-1}, t_k]$ for $1 \leq k \leq m$.

Suppose $Y_{i,1}, ..., Y_{i,k_c}$ $(i = 1, 2...n)$ are i.i.d. random variables that follow distribution function $F(\theta_1)$, while $Y_{i,k_c+1}, ...Y_{i,m}$ $(i = 1, 2...n)$ are i.i.d. random variables that follow distribution function $F(\theta_2)$. If $\theta_1 \neq \theta_2$, then t_{k_1+1} is the only changepoint. Otherwise, there is no changepoint. Without loss of generosity, the prior on changepoint probability is assumed to be p uniformly.

As derived in the previous chapter, we have the posterior changepoint probability as follows,

$$\widehat{p}_{k+1} = \frac{p}{p + \sum_{1 \leq j \leq k < l \leq m} (1-p) p_{jk} q_{k+1,l} f_{jl} / (f_{jk} f_{k+1,l})}. \tag{3.43}$$

where

$$p_{jk} \propto p^*_{jk} = \begin{cases} p_k f_{kk} & j = k, \\ (1 - p_k) p_{j,k-1} \dfrac{f_{jk}}{f_{j,k-1}} & j < k, \end{cases} \tag{3.44}$$

$$q_{k+1,l} \propto q^*_{k+1,l} = \begin{cases} p_{k+1} f_{k+1,k+1} & l = k+1, \\ (1 - p_{k+1}) q_{k+2,l} \dfrac{f_{k+1,l}}{f_{k+2,l}} & l > k+1, \end{cases} \tag{3.45}$$

and

$$f_{jk} = f(Y_{jk}|C_{jk}) = \int_{-\infty}^{\infty} f(Y_{jk}|\theta)\Pi(\theta)d\theta \tag{3.46}$$

Lemma 1. If $I_j = I_{j+1} = \ldots = I_l = 0$, then as $n \to \infty$, $\frac{f_{jl}}{f_{jk} f_{k+1,l}} \to \infty$, for all $j,\ k,\ l$ where $j \leq k < l$.

The interpretation of this lemma is that if no changepoint is present between j and l, then for any k between j and l, the ratio between the density f_{jl} based on a uniform θ and the density product $f_{jk}f_{k+1,l}$ based on two different θ's will converge to infinity as the number of observations increases.

Proof of Lemma 1

$$
\begin{aligned}
f_{jk} &= \int_{-\infty}^{\infty} \prod_{\alpha=j}^{k} \prod_{i=1}^{n} f_\theta(Y_{i\alpha}) \Pi(\theta) d\theta \\
&= \int_{-\infty}^{\infty} e^{n(k-j+1) \cdot \frac{1}{n(k-j+1)} \sum_{i,\alpha} log f_\theta(Y_{i\alpha})} \Pi(\theta) d\theta
\end{aligned}
$$

Using $f_{jk}(\theta)$ to stand for the likelihood function $\prod_{\alpha=j}^{k} \prod_{i=1}^{n} f_\theta(Y_{i\alpha})$ and expanding the above expression at the point of the corresponding MLE $\hat{\theta}_{jk}$ yields

$$
f_{jk} \doteq \frac{\sqrt{2\pi}\Pi(\hat{\theta}_{jk})}{\sqrt{-L''(\hat{\theta}_{jk})}} \cdot \frac{f_{jk}(\hat{\theta}_{jk})}{\sqrt{n(k-j+1)}}. \tag{3.47}
$$

Likewise,

$$
f_{k+1,l} \doteq \frac{\sqrt{2\pi}\Pi(\hat{\theta}_{k+1,l})}{\sqrt{-L''(\hat{\theta}_{k+1,l})}} \cdot \frac{f_{k+1,l}(\hat{\theta}_{k+1,l})}{\sqrt{n(l-k)}}, \tag{3.48}
$$

and

$$
f_{jl} \doteq \frac{\sqrt{2\pi}\Pi(\hat{\theta}_{jl})}{\sqrt{-L''(\hat{\theta}_{jl})}} \cdot \frac{f_{jl}(\hat{\theta}_{jl})}{\sqrt{n(l-j+1)}}. \tag{3.49}
$$

Combining the three terms together, we get

$$
\frac{f_{jl}}{f_{jk}f_{k+1,l}} \sim \sqrt{n} e^{\log(f_{jk}(\hat{\theta}_{jl})) - \log(f_{jk}(\hat{\theta}_{jk}))} \cdot e^{\log(f_{k+1,l}(\hat{\theta}_{jl})) - \log(f_{k+1,l}(\hat{\theta}_{k+1,l}))}. \tag{3.50}
$$

Given no changepoint between j and l, $\log(f_{jk}(\hat{\theta}_{jl})) - \log(f_{jk}(\hat{\theta}_{jk})) \sim O(1)$. As a result,

$$\frac{f_{jl}}{f_{jk}f_{k+1,l}} = O(\sqrt{n}), \quad \text{when } n \to \infty. \tag{3.51}$$

Lemma 2. If $I_j = I_{j+1} = \ldots = I_k = 0$, $I_{k+1} = 1$ and $I_{k+2} = \ldots = I_l = 0$ then as $n \to \infty$, $\frac{f_{jl}}{f_{jk}f_{k+1,l}} \to 0$, for all $j,\ l$ where $j \leq k < l$.

Lemma 2 is the opposite of **Lemma 1**. It states that if $k+1$ is the only changepoint between j and l, then the ratio between the density f_{jl} based on a uniform θ and the density product $f_{jk}f_{k+1,l}$ based on the correct changepoint of θ will converge to zero as the number of observations increases.

Proof of Lemma 2

Following the same steps of proving **Lemma 1**,

$$\frac{f_{jl}}{f_{jk}f_{k+1,l}} \sim \sqrt{n}e^{\log(f_{jk}(\hat{\theta}_{jl}))-\log(f_{jk}(\hat{\theta}_{jk}))} \cdot e^{\log(f_{k+1,l}(\hat{\theta}_{jl}))-\log(f_{k+1,l}(\hat{\theta}_{k+1,l}))}. \tag{3.52}$$

Because $k+1$ is the only changepoint between j and l,

$$\log(f_{jk}(\hat{\theta}_{jl})) - \log(f_{jk}(\hat{\theta}_{jk})) < 0$$

when $n \to \infty$. As a result,

$$\frac{f_{jl}}{f_{jk}f_{k+1,l}} = O(\sqrt{n}e^{-n}) \to 0, \quad \text{when } n \to \infty. \tag{3.53}$$

Lemma 3. If $I_j = 1$ and $I_{j+1} = \ldots = I_k = 0$ then as $n \to \infty$, $p_{jk} \to 1$, for all $j,\ k$ where $j \leq k$.

Lemma 3 states that if j is a changepoint and there is no other changepoint between j and k, then the probability of j being the most recent changepoint of k converges to 1 as the number of observations increases.

Proof of Lemma 3

(a) Because k_c is the only changepoint, based on **Lemma 2** we have

$$\frac{f_{j,k_c+1}}{f_{jk_c} f_{k_c+1,k_c+1}} \to 0, \quad \text{for all } j \leq k_c. \tag{3.54}$$

Inserting this into the expression of p_{k_c+1,k_c+1}, we get

$$p_{k_c+1,k_c+1} = \frac{p}{p + (1-p)\sum_{j \leq k_c} p_{j,k_c} \frac{f_{j,k_c+1}}{f_{j,k_c} f_{k_c+1,k_c+1}}} \to 1 \tag{3.55}$$

when $n \to \infty$.

(b) Assume we have $p_{k_c+1,k} \to 1$ for $k \geq k_c{+}1$, we would like to prove $p_{k_c+1,k+1} \to 1$. Given $p_{k_c+1,k} \to 1$ and $\sum_{j<=k} p_{jk} = 1$, we have

$$p_{jk} \to 0, \quad \text{for all } j \neq k_c + 1. \tag{3.56}$$

Also,

$$\begin{aligned} p_{k_c+1,k+1} &= \frac{(1-p)p_{k_c+1,k}\frac{f_{k_c+1,k+1}}{f_{k_c+1,k}}}{p f_{k+1,k+1} + (1-p)\sum_{j \leq k} p_{jk}\frac{f_{j,k+1}}{f_{jk}}} \\ &= \frac{1-p}{p\frac{f_{k_c+1,k} f_{k+1,k+1}}{f_{k_c+1,k+1}} + (1-p) + (1-p)\sum_{j \leq k_c} p_{jk} A_{jk} + \sum_{k_c+1<j\leq k} p_{jk} A_{jk}}, \end{aligned} \tag{3.57}$$

where

$$A_{jk} = \frac{f_{j,k+1} f_{k_c+1,k}}{f_{jk} f_{k_c+1,k+1}} = \frac{f_{j,k+1}}{f_{jk} f_{k+1,k+1}} \cdot \frac{f_{k_c+1,k} f_{k+1,k+1}}{f_{k_c+1,k+1}} \tag{3.58}$$

According to **Lemma 2**,

$$\frac{f_{k_c+1,k} f_{k+1,k+1}}{f_{k_c+1,k+1}} \to 0, \tag{3.59}$$

$$A_{jk} \sim O(1), \quad \text{for all } j > k_c + 1 \tag{3.60}$$

and

$$A_{jk} \to 0, \quad \text{for all } j < k_c + 1. \tag{3.61}$$

Combining (3.56), (3.58), (3.59), (3.60), (3.61) with (3.57) yields $p_{k_c+1,k+1} \to 1$. Combining (a) and (b), we have proven **Lemma 3**.

Symmetrically, the same property holds for the very next changepoint: If $l+1$ is a changepoint and no other changepoint exists between $k+1$ and l, then the probability of $l+1$ being the most recent changepoint of $k+1$ converges to 1 as the number of observations increases.

Lemma 4. If $\widetilde{I}_l = 1$ and $\widetilde{I}_k = \ldots = \widetilde{I}_{l-1} = 0$ then as $n \to \infty$, $q_{k+1,l} \to 1$, for all $k,\ l$ where $l \geq k+1$.

Following **Lemma 3** and also the facts that $\sum_{j=1}^{k} p_{jk} = 1$, we get the corollary that if there exists a changepoint before k and $j \leq k$ is not a changepoint, then the probability of j being the most recent changepoint of k converges to 0 as the number of observations increases.

Corollary 1 If $I_{j'} = 1$ and $I_j = 0$, then as $n \to \infty$, $p_{jk} \to 0$, for all $j,\ k$ where $j \leq k$ and $j' \leq k$.

Likewise, if there exists a changepoint after k and $l+1$ is not a changepoint, then the probability of $l+1$ being the very next changepoint of k converges to 0 as the number of observations increases.

Corollary 2 If $\widetilde{I}_{l'} = 1$ and $\widetilde{I}_l = 0$, then as $n \to \infty$, $q_{k+1,l} \to 0$, for all $k,\ l$ where $l \geq k+1$ and $l' \geq k+1$.

3.3.1.1 No Changepoint

Theorem 1. If $I_2 = \ldots = I_m = 0$, then as $n \to \infty$, $\hat{p}_{k+1} \to 0$ for all $k = 1, \ldots, m-1$.

Proof of Theorem 1

Using **Lemma 1**, we have

$$\frac{f_{jl}}{f_{jk} f_{k+1,l}} \to \infty,$$

for all $j,\ k,\ l$ where $1 \leq j \leq k < l \leq m$. Combining this with (3.43) yields $\hat{p}_{k+1} \to 0$ for all $k = 1, \ldots, m-1$.

3.3.1.2 One Changepoint

Theorem 2. If $I_2 = \ldots = I_{k_c} = 0$, $I_{k_c+1} = 1$ and $I_{k_c+2} = \ldots = I_m = 0$, then as $n \to \infty$, (1) $\hat{p}_{k_c+1} \to 1$; (2) $\hat{p}_{k+1} \to 0$ for all $k = 1, \ldots, m-1$ and $k \neq k_c$.

Theorem 2 states the fact that if there is one single changepoint $k_c + 1$, the posterior changepoint probability $\hat{p}_{k+1}$ on the changepoint converges to 1 while the posterior changepoint probability on other time points converges to zero.

Proof of Theorem 2

(1) Using **Lemma 2** and the fact that k_c is the only changepoint, we get

$$\frac{f_{jl}}{f_{jk_c} f_{k_c+1,l}} \to 0,$$

for all $j,\ l$ where $j \leq k < l$. Combining this with (3.43) yields $\hat{p}_{k_c+1} \to 1$.

(2) Assume $k > k_c$ without loss of generality. Given the fact that there is no changepoint between $k_c + 1$ and l, according to **Lemma 1**,

$$\frac{f_{k_c+1,l}}{f_{k_c+1,k} f_{k+1,l}} \to \infty \tag{3.62}$$

for all $l > k$.

Given the fact that $k_c + 1$ is indeed the most recent changepoint prior to k and according to **Lemma 3**, we get

$$p_{k_c+1,k} \to 1. \tag{3.63}$$

Combining (3.62) and (3.63) yields

$$\sum_{j,l} \frac{f_{jl}}{f_{jk} f_{k+1,l}} p_{jk} q_{k+1,l} > p_{k_c+1,k} \sum_{l} \frac{f_{k_c+1,l}}{f_{k_c+1,k} f_{k+1,l}} q_{k+1,l} \to \infty. \tag{3.64}$$

Plugging the above result into (3.43) yields $\hat{p}_{k+1} \to 0$.

Symmetrically, the same argument can be made for $k < k_c$ using **Lemma 4**.

3.4 Simulation Studies

In this section, some simulation studies are done to show the performance of the proposed Cox model changepoint method under a number of different scenarios. A test of robustness is also performed to see how the method works when the actual change is gradual instead of abrupt.

In general, the failure times are generated using the following property:

Property. If T is a failure time whose cumulative hazard function is $\Lambda(\cdot)$, then the random variable $\Lambda(T)$ follows the exponential distribution with parameter 1.

Hence, any failure time with a known hazard function, as in the Cox model, can be generated by applying the inverse cumulative hazard function on an exp(1) random variable.

3.4.1 Fixed Segmentation

In this subsection, we consider the case where $\beta(t)$ has one single changepoint. We examine whether the changepoint procedure could consistently identify the correct changepoint location. The fixed segmentation method is used in this subsection with the length of segments being 0.1.

Unless otherwise specified, all simulation results reported in this section are based on 1000 replications.

In the first series of simulations, the single covariate X is generated from a

Table 3.1: One changepoint with Bernoulli covariate, fixed segmentation.

Sample Size	Effective Sample Size	Correct	Incorrect
Constant baseline hazard, exponential censoring			
100	59	592	408
300	176	863	137
500	293	905	95
Constant baseline hazard, uniform censoring			
100	67	806	194
300	202	917	83
500	337	951	49
Linear baseline hazard, exponential censoring			
100	68	759	241
300	204	943	57
500	339	972	28
Linear baseline hazard, uniform censoring			
100	89	894	106
300	267	960	40
500	444	985	15

Bernoulli distribution with $p = 0.5$. The changepoint is at $t = 4$ with $\beta(t) = 0$ if $t \leq 4$ and $\beta(t) = 2$ if $t > 4$. Three different sample sizes are considered: a) 100, b) 300 and c) 500. Two types of baseline hazard function are included: a) constant $\lambda_0(t) = 0.1$ and b) linear with respect to time $\lambda_0(t) = 0.087t$. Two types of censoring are considered: a) exponential with parameter 10, $C \sim \exp(10)$ and b) uniform between 3 and 8, $C \sim \text{unif}(3, 8)$. The combinations result in a total of 12 cases considered.

Table 3.1 reports the "effective sample size", meaning the number of failure times excluding censored observations, the number of replications in which the changepoint estimate is in the correct segment and the number of replications the

Table 3.2: One changepoint with normal covariate, fixed segmentation.

Sample Size	Effective Sample Size	Correct	Incorrect
Constant baseline hazard, exponential censoring			
100	50	735	265
300	150	953	47
500	250	990	10
Constant baseline hazard, uniform censoring			
100	56	853	147
300	169	986	14
500	281	996	4
Linear baseline hazard, exponential censoring			
100	62	882	118
300	187	989	11
500	311	1000	0
Linear baseline hazard, uniform censoring			
100	76	962	38
300	229	1000	0
500	382	1000	0

estimate is in an incorrect segment. It can be observed that when sample size is only 100 and heavy censoring exists, the percentage of the procedure correctly identifying the changepoint segment can be as low as 60%. Under milder censoring, or when sample size is 500, the percentage of correct identification rises to 95% and above.

In the second series of simulations, we consider a covariate X following the standard normal distribution. The changepoint is still at 4 and β still changes from 0 to 2. The combinations of sample size, baseline hazard and censoring are also kept the same.

The results in Table 3.2 are quite similar to Table 3.1. The identification power of the procedure is even better under a normal covariate, hence increasing the percentage of correct identification to close to 100% in some cases.

3.4.2 One Changepoint

In this subsection, we consider the case where the true $\beta(t)$ has one single changepoint. We examine whether the changepoint procedure could consistently identify the correct changepoint location and consequently the effect β before and after the changepoint. The natural segmentation is used in this subsection, meaning the time axis is segmented by observed failure times.

3.4.2.1 Number of Changepoints Known

We first examine the scenario where the number of changepoints is known. In this case, the procedure is obviously to identify the time point with the greatest posterior changepoint probability as the changepoint estimate.

In the first series of simulations, the single covariate X is generated from a Bernoulli distribution with $p = 0.5$. The changepoint is at $t = 4$ with $\beta(t) = 0$ if $t \leq 4$ and $\beta(t) = 2$ if $t > 4$. Three different sample sizes are considered: a) 100, b) 300 and c) 500. Two types of baseline hazard function are included: a) constant $\lambda_0(t) = 0.1$ and b) linear with respect to time $\lambda_0(t) = 0.087t$. Two types of censoring are considered: a) exponential with parameter 10, $C \sim \exp(10)$ and b) uniform between 3 and 8, $C \sim \text{unif}(3, 8)$. The combinations result in a total of 12 cases considered.

Table 3.3: One changepoint with Bernoulli covariate, natural segmentation.

Sample	Effective	Changepoint		$\hat{\beta}_1$		$\hat{\beta}_2$	
Size	Sample Size	Bias	RMSE	Bias	RMSE	Bias	RMSE
Constant baseline hazard, exponential censoring							
100	59	3.3148	8.1935	0.0259	0.6604	0.5786	6.2543
300	176	-0.0629	1.2555	-0.0292	0.2376	-0.0174	0.3040
500	293	-0.0764	0.2155	-0.0098	0.1747	-0.0111	0.2116
Constant baseline hazard, uniform censoring							
100	67	-0.3143	0.8546	-0.0969	0.4360	0.1678	3.0752
300	202	-0.1275	0.3180	-0.0261	0.2140	-0.0217	0.2664
500	337	-0.0611	0.1640	-0.0111	0.1574	-0.0123	0.1907
Linear baseline hazard, exponential censoring							
100	68	0.6395	2.4251	-0.0254	0.5425	0.1122	4.3991
300	204	-0.0700	0.1931	-0.0243	0.1992	-0.0294	0.3107
500	339	-0.0392	0.1050	-0.0212	0.1581	-0.0060	0.2184
Linear baseline hazard, uniform censoring							
100	89	-0.1628	0.6218	-0.0764	0.3562	-0.0278	1.3349
300	267	-0.0461	0.1407	-0.0140	0.1792	-0.0205	0.2269
500	444	-0.0255	0.0728	-0.0174	0.1349	0.0105	0.1754

Table 3.3 reports the effective sample size, bias and root mean square error (RMSE) of changepoint and β estimates of each case considered. When the sample size is small (100), the effective sample size is even smaller because of censoring, it can be seen that there are certain replications out of the total of 1000 that produce results that are outliers, which result in relatively large bias and RMSE reported. In other words, under a small sample size, there could be situations that the changepoint cannot be identified or correctly identified. There is a significant improvement when the sample size increases to 300 and even better result when it goes to 500. On the other hand, the method seems to work well under both baseline hazard functions and both types of censoring.

In the second series of simulations, we consider a covariate X following the standard normal distribution. The changepoint is still at 4 and β still changes from 0 to 2. The combinations of sample size, baseline hazard and censoring are also kept the same.

The changepoint method is shown to work well in Table 3.4. Outliers can still be observed when sample size is at 100, but with less frequency, as evidenced by smaller bias and RMSE. Also in general, bias and RMSE reported in Table 3.4 are noticeably better compared to their counterparts in Table 3.3. This is because the covariate in Table 3.4, being normally distributed, has a much wider dispersion than the Bernoulli covariate in Table 3.3.

In the third series of simulations, different true changepoint locations are tested

Table 3.4: One changepoint with normal covariate, natural segmentation.

Sample Size	Effective Sample Size	Changepoint Bias	Changepoint RMSE	$\hat{\beta}_1$ Bias	$\hat{\beta}_1$ RMSE	$\hat{\beta}_2$ Bias	$\hat{\beta}_2$ RMSE
Constant baseline hazard, exponential censoring							
100	50	0.5149	4.3480	-0.0090	0.2420	0.8870	12.7320
300	150	-0.0117	0.0813	-0.0043	0.1108	-0.0655	0.2280
500	250	-0.0083	0.0545	-0.0044	0.0879	0.0184	0.1689
Constant baseline hazard, uniform censoring							
100	56	-0.0477	0.2636	-0.0126	0.1933	-0.1744	0.3684
300	169	-0.0114	0.0654	-0.0023	0.1024	-0.0642	0.1987
500	281	-0.0046	0.0331	0.0017	0.0806	-0.0371	0.1528
Linear baseline hazard, exponential censoring							
100	62	0.2542	2.6838	-0.0099	0.1954	0.7323	11.9094
300	187	-0.0070	0.0397	-0.0051	0.1013	-0.0630	0.2137
500	311	-0.0027	0.0194	-0.0026	0.0757	-0.0482	0.1625
Linear baseline hazard, uniform censoring							
100	76	-0.0268	0.1269	-0.0078	0.1564	-0.1525	0.3494
300	229	-0.0034	0.0224	0.0027	0.0896	-0.0480	0.1910
500	382	-0.0027	0.0155	0.0037	0.0664	-0.0342	0.1415

Table 3.5: Varying changepoint locations.

Changepoint Location	Effective Sample Size	Changepoint Estimate Bias	RMSE	$\hat{\beta}_1$ Bias	RMSE	$\hat{\beta}_2$ Bias	RMSE
4	250	-0.0083	0.0545	-0.0044	0.0879	0.0184	0.1689
2	250	-0.0040	0.0331	-0.0040	0.1159	0.0069	0.1394
8	250	-0.0201	0.1178	-0.0040	0.0731	0.0372	0.2700

and compared against each other. Three different locations are used: a) 4, b) 2, c) 8. In all cases, β changes from 0 to 2. Sample size is fixed at 500 here and the covariate is normally distributed. The baseline hazard rate is constant and censoring follows the exp(10) distribution.

Since the only difference between the three cases is the location of the changepoint, one can see in Table 3.5 that they share the same effective sample size. In other words, about 50% of observations are censored in each case. In terms of the estimation of β_1, RMSE is smallest when the changepoint is at time 8 and largest when changepoint is at 2. Conversely, the RMSE of β_2 is smallest when the changepoint is at time 2 and largest at time 8. Both of these are intuitive trends since the more observations one has for a segment, the better the estimation of the parameter will be. The interesting fact to notice is that the RMSE of the changepoint estimate is smallest when the changepoint is at time 2, which coincides with the case where the RMSE's of β_1 and β_2 are the closest to each other among the three cases studied. Conceptually, this suggests that a changepoint can be best

Table 3.6: Varying degrees of censoring.

	Effective	Changepoint Estimate		$\hat{\beta}_1$		$\hat{\beta}_2$	
Censoring	Sample Size	Bias	RMSE	Bias	RMSE	Bias	RMSE
exp(10)	250	-0.0083	0.0545	-0.0044	0.0879	0.0184	0.1689
exp(5)	176	-0.0180	0.0918	-0.0029	0.0949	-0.0701	0.2269
exp(20)	316	-0.0050	0.0392	-0.0012	0.0844	-0.0249	0.1363

identified when there is about equal information of the parameter before and after the change.

In the fourth series of simulations, different censoring distributions are tested and compared against each other. The three cases studied are all exponential distribution with different parameters: a) 10, b) 5, c) 20, hence case b) has the heaviest censoring, case c) the lightest and a) in the middle. Sample size is fixed at 500 here and the covariate is normally distributed. The baseline hazard rate is constant and the changepoint is at time 4 with β jumping from 0 to 2.

As displayed in Table 3.6 we have an effective sample size of 250 in case a), 176 in case b) and 316 in case c), in the order expected. The lighter the censoring, the better the estimations of the changepoint, β_1 and β_2 are, which is consistent with intuition. However, the impact of censoring on $\hat{\beta}_1$ is only marginal compared to $\hat{\beta}_2$. This is because heavier censoring decreases observed failures after the changepoint much more than before the changepoint. For example, in case a), on average there are 137 observations before the changepoint and 112 after. Under heavier censoring

Table 3.7: Varying scales of change

True β_2	Effective Sample Size	Changepoint Estimate Bias	Changepoint Estimate RMSE	$\hat{\beta}_1$ Bias	$\hat{\beta}_1$ RMSE	$\hat{\beta}_2$ Bias	$\hat{\beta}_2$ RMSE
Sample size = 100							
2	50	0.5149	4.3480	-0.0090	0.2420	0.8870	12.7320
8	50	-0.0504	0.1867	-0.0144	0.2068	0.2368	2.2915
Sample size = 300							
2	150	-0.0117	0.0813	-0.0043	0.1108	-0.0655	0.2280
8	150	-0.0088	0.0365	-0.0053	0.1099	0.0218	0.9043

of case b), the numbers become 116 and 60, respectively. The percentage of decrease is about 15% before the changepoint and more than 45% after the changepoint.

The fifth series of simulations are concentrated on different scales of change in β. We study 4 cases, the first pair with sample size 100 and the second pair 300. Each pair will be a contrast of one case with β changing from 0 to 2 and another with β changing from 0 to 8. The changepoint location is fixed at 4 in all cases. Intuitively, sharper change in the parameter of interest should lead to better identification of the changepoint. The covariate is normally distributed, and the baseline hazard rate is constant.

Consistent with our intuition, it can be seen in Table 3.7 that the changepoint estimate improves when β's scale of change becomes greater, regardless of sample size. The improvement when sample size is only 100 is more noticeable. The greater scale of change decreases the number of outlier replications. As a result, the changepoint estimation improves significantly.

Table 3.8: One changepoint with number of changepoint unknown.

Sample Size	1 cp	0 cp	≥ 2 cp	Total
Constant baseline hazard, exponential censoring				
100	701	156	143	1000
300	803	71	126	1000
500	866	58	76	1000
Constant baseline hazard, uniform censoring				
100	768	92	140	1000
300	852	61	87	1000
500	887	63	50	1000

Note. cp stands for changepoint.

3.4.2.2 Number of Changepoints Unknown

In contrast to previous subsections, in this subsection, we assume that the number of changepoints is unknown in the simulations while keeping other elements the same. In order to identify changepoints in this setup, we have to choose a threshold parameter $\hat{p}_T$ and an accuracy parameter τ. Every time point with the posterior changepoint probability greater than $\hat{p}_T$ will be identified as potential changepoints. All potential changepoints then are rank ordered in descendance and any points within a distance τ to a higher ranked changepoints are eliminated. The parameters chosen here are $\hat{p}_T = 0.25$ and $\tau = 0.05$.

Table 3.8 reports the number of replications out of the total of 1000 in which the changepoint procedure identifies 1 changepoint correctly, no changepoint and greater than or equal to 2 changepoints, respectively. As the sample size increases

Table 3.9: Zero changepoint with number of changepoint unknown.

Sample Size	0 cp	≥ 1 cp	Total
	$\beta = 0$		
100	1000	0	1000
300	1000	0	1000
500	999	1	1000
	$\beta = 2$		
100	1000	0	1000
300	1000	0	1000
500	999	1	1000

Note. cp stands for changepoint.

from 100 to 500, the percentage of correct number of changepoints identified increases from 70% to above 85%.

For comparison and benchmarking the effectiveness of the procedure, we also performed another series of simulations where the true $\beta(t)$ is constant with respect to time. Ideally, the changepoint procedure should identify zero changepoint in most situations, thus minimizing what can be regarded as the "type I" error rate. Two cases are studied here with a) $\beta = 0$ in one and b) $\beta = 2$ in the other. The same parameters $\hat{p}_T = 0.25$ and $\tau = 0.05$ are used for consistency.

Table 3.9 shows that the procedure indeed performs as expected. The chance of the procedure identifying a changepoint when the true effect has no changepoint is indeed minimal in all scenarios studied.

3.4.3 Two Changepoints

In this subsection, we consider the case where the true $\beta(t)$ undergoes two jumps. We examine whether the changepoint procedure could consistently identify the correct changepoint location and consequently the effect β before and after the changepoint. The natural segmentation is used in this subsection, meaning the time axis is segmented by observed failure times.

3.4.3.1 Number of Changepoints Known

In the first scenario, we assume the number of changepoints, 2, is known. Hence, the procedure used is similar to when the number of changepoints is known to be 1. The two time points with the highest posterior changepoint probability are identified as changepoints provided that they are apart by a distance greater than $\tau = 0.05$.

In this series of simulations, we study two cases: a) constant baseline hazard, exponential censoring, $\beta(t)$ is piecewise constant (0, -2, 0) with changepoints at time 2 and 10; b) Linear baseline hazard, exponential censoring, $\beta(t)$ is piecewise constant (0, -2, 0) with changepoints at time 3 and 7.

Table 3.10 reports, out of 1000 replications, the number of replications where cp1 and cp2 are each correctly identified (defined as the estimate in the neighborhood of the true changepoint), as well as the number of replications where they are both correctly identified. It can be seen that the percentage of simultaneous correct

Table 3.10: Two changepoints with number of changepoint known (a).

Sample Size	cp1	cp2	Both
Constant baseline hazard, exponential censoring			
300	979	902	881
500	975	960	935
Linear baseline hazard, exponential censoring			
300	977	928	905
500	984	972	956

Note. cp stands for changepoint.

Table 3.11: Two changepoints with number of changepoint known (b).

		cp1		cp2	
Sample Size	Success Rate	Bias	RMSE	Bias	RMSE
Constant baseline hazard, exponential censoring					
300	88.1%	0.0078	0.0615	0.0840	0.6742
500	93.5%	0.0040	0.0322	0.0239	0.3468
Linear baseline hazard, exponential censoring					
300	90.5%	0.0074	0.0473	0.0101	0.2052
500	95.6%	0.0035	0.0267	-0.0242	0.1680

Note. CP stands for changepoint.

identification is around 90% when the sample size is 300 and rises to about 95% when the sample size is 500.

Table 3.11 reports the success rate (percentage of simultaneous correct identification) and the bias and RMSE of the two changepoint estimates across all successful replications.

Table 3.12: Two changepoints with number of changepoint unknown (a).

Sample Size	2 cp	0 cp	1 cp	>2 cp	Total
Constant baseline hazard, exponential censoring					
300	536	8	202	254	1000
500	599	7	180	214	1000
Linear baseline hazard, exponential censoring					
300	582	12	209	197	1000
500	636	6	204	154	1000

Note. cp stands for changepoint.

3.4.3.2 Number of Changepoints Unknown

In the first scenario, we assume the number of changepoints is unknown. Hence, the procedure used is the same as when the true number of changepoints, 1, is unknown as in Section 3.4.2.2. The same parameters are also used, $\hat{p}_T = 0.25$ and $\tau = 0.05$.

In this series of simulations, we study two cases: a) constant baseline hazard, exponential censoring, $\beta(t)$ is piecewise constant (0, -2, 0) with changepoints at time 2 and 10; b) Linear baseline hazard, exponential censoring, $\beta(t)$ is piecewise constant (0, -2, 0) with changepoints at time 3 and 7.

Table 3.12 reports, out of 1000 replications, the number of replications where the number of changepoints identified is 2 (correct), 0 or 1 (too few) and more than 2 (too many). It can be seen that the percentage of simultaneous correct identification is around 55% when the sample size is 300 and rises to about 60% when the sample size is 500.

Table 3.13: Two changepoints with number of changepoint unknown (b).

		cp1		cp2	
Sample Size	Success Rate	Bias	RMSE	Bias	RMSE
Constant baseline hazard, exponential censoring					
300	49.8%	0.0094	0.0565	0.0873	0.4748
500	57.8%	0.0041	0.0333	0.0385	0.3060
Linear baseline hazard, exponential censoring					
300	53.7%	0.0086	0.0428	0.0144	0.1797
500	62.9%	0.0044	0.0236	-0.0144	0.1510

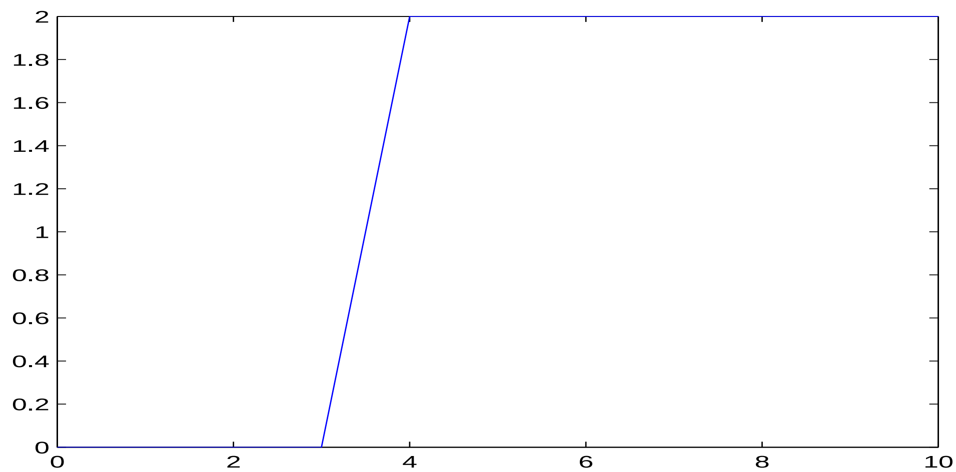
Note. cp stands for changepoint.

Table 3.13 reports the success rate (percentage of simultaneous correct identification) and the bias and RMSE of the two changepoint estimates across all successful replications. One will notice that the success rates in Table 3.13 are a bit lower than those in Table 3.12. This is because there are replications where the correct number of changepoints, 2, is identified but the locations of the two changepoints are incorrect. A typical example of misidentification would be two estimated changepoints being both around 7 while the true changepoints are 3 and 7.

3.4.4 A Study of Robustness

So far we have performed simulations under the assumption that the true $\beta(t)$ does undergo an abrupt change, or a jump. In practice, it is not unreasonable to assume that $\beta(t)$ actually goes through a continuous or gradual change. For example, Martinussen and Scheike (2006) studied extensively such models with

Figure 3.1: Treatment effect steep change



time-varying effects of covariates. Therefore, it would be of interest to test how the proposed changepoint method performs under such continuous changes.

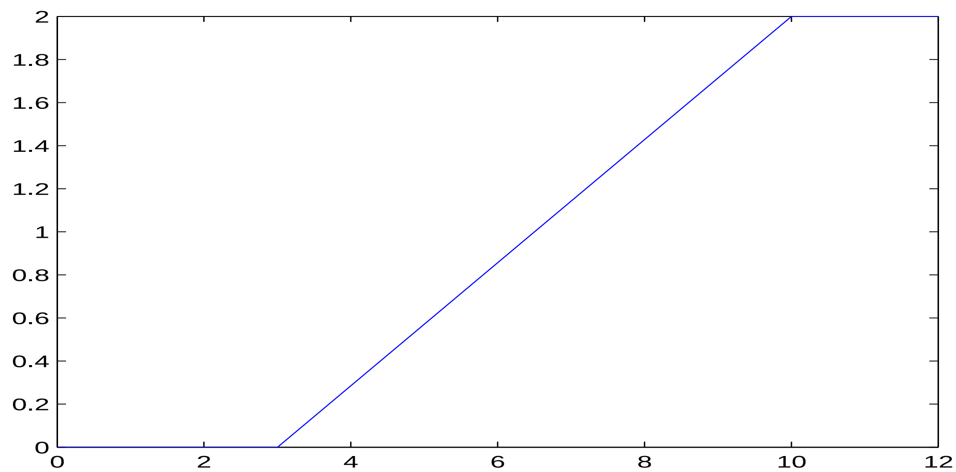
First, we examine a scenario where $\beta(t)$ undergoes a continuous but steep linear change between time 3 and 4 and then stays constant as in Figure 3.1. Namely,

$$\beta(t) = \begin{cases} 0 \quad \text{if } 0 \leq t < 3, \\ 2(t-3) \quad \text{if } 3 \leq t < 4, \\ 2 \quad \text{if } t \geq 4. \end{cases}$$

In the second scenario, $\beta(t)$ undergoes a continuous and gradual change between time 3 and 10 and then stays constant as in Figure 3.2. Namely,

$$\beta(t) = \begin{cases} 0 \quad \text{if } 0 \leq t < 3, \\ \frac{2}{7}(t-3) \quad \text{if } 3 \leq t < 10, \\ 2 \quad \text{if } t \geq 10. \end{cases}$$

Figure 3.2: Treatment effect gradual change



Results of the changepoint estimation are displayed in Table 3.14. In the "steep change" case, the mean of changepoint estimates is 3.4562 and right between 3 and 4. In fact, out of all 1000 replications, 98.5% of them returned a changepoint estimate in the range of $[3, 4]$. In the "gradual change" case, the mean of estimated changepoint is also between 3 and 10. Similarly, 99.4% of all replications returned a changepoint estimate in the range of $[3, 10]$. It is worthwhile to note though that the standard deviation of estimates is 1.2251, much larger than the standard deviation in the "steep change" case.

To summarize, we have shown that even if $\beta(t)$ goes through a continuous change, the proposed changepoint method still produces a desirable estimate that falls within the range of β change around 99% of the time.

Table 3.14: Treatment effect continuous change.

Changepoint Mean	Estimate SD	Percentage of estimate falling in between
β steep change between $[3, 4]$		
3.4562	0.1845	98.50 %
β gradual change between $[3, 10]$		
5.4712	1.2251	99.40 %

Note. SD stands for standard deviation.

3.5 Data Analysis

The real data we analyze in this section is a randomized clinical trial as in Stablein and Koutrouvelis (1985). The clinical trial followed the survival of two groups of gastric cancer patients, one group receiving chemotherapy treatment only and the other receiving both chemotherapy and radiotherapy.

Figure 3.3 displays the Kaplan-Meier estimates of survival probabilities for the two groups of gastric cancer patients. The crossing of the two survival curves is a strong indication of crossing hazards, and strongly suggests the possibility of a changepoint in the hazard of our model.

Both our proposed changepoint procedure and Luo and Turnbull's MLE procedure are applied to the clinical trial data. Interestingly, they yield basically the same result: a changepoint is found on the 31st failure time, meaning 262 days. Figure 3.4 and Figure 3.5 display the result of the two procedures, respectively. The β estimates before and after the changepoint are 1.4229 and -0.64228 respectively,

Figure 3.3: Kaplan-Meier estimates of survival probabilities

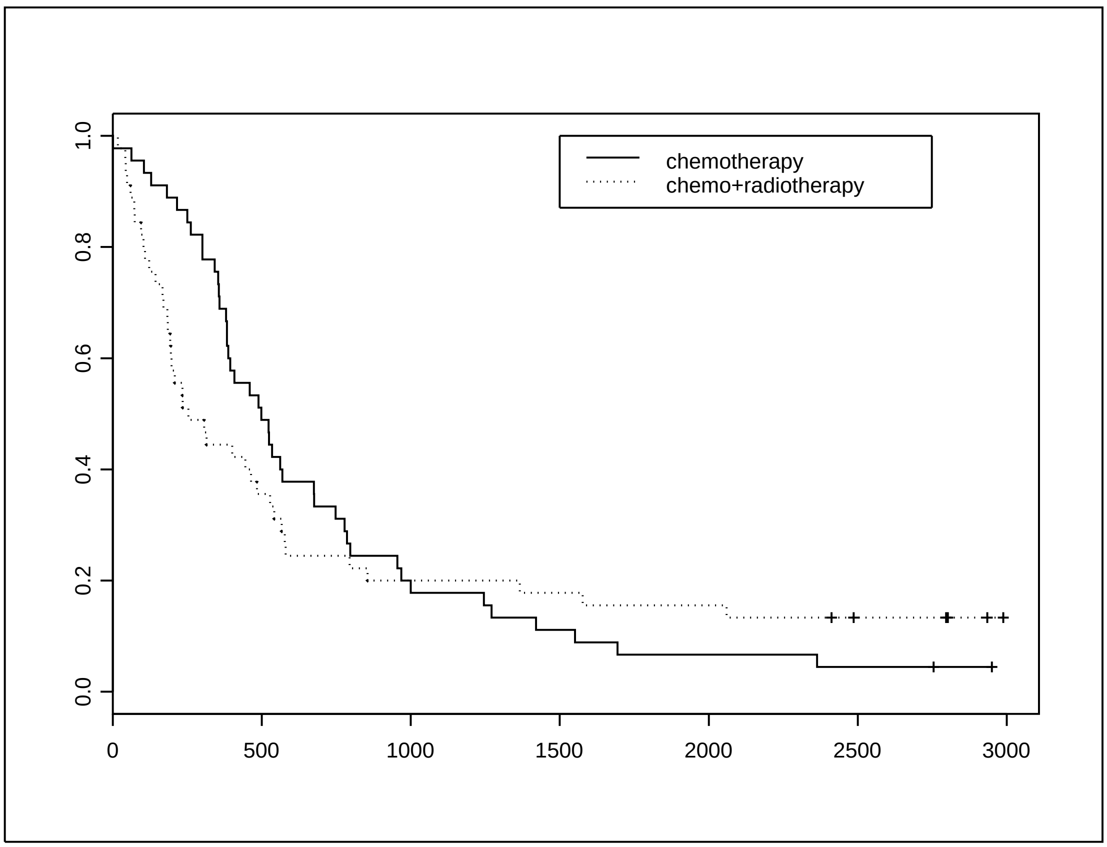
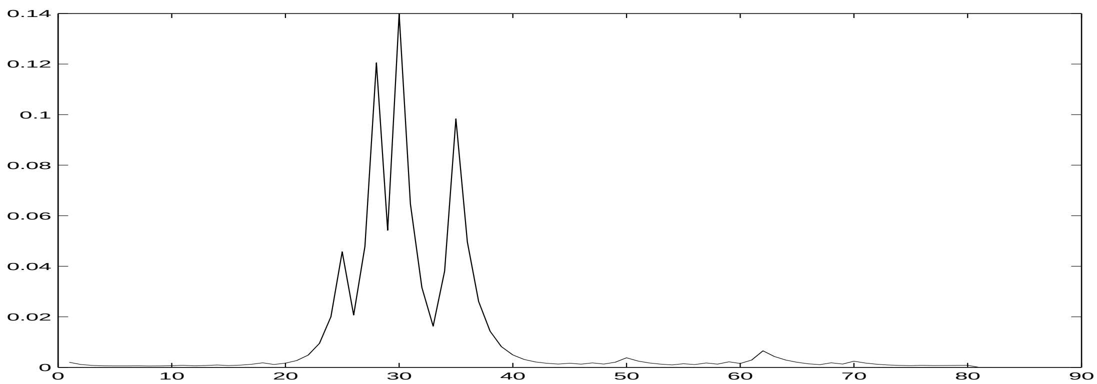


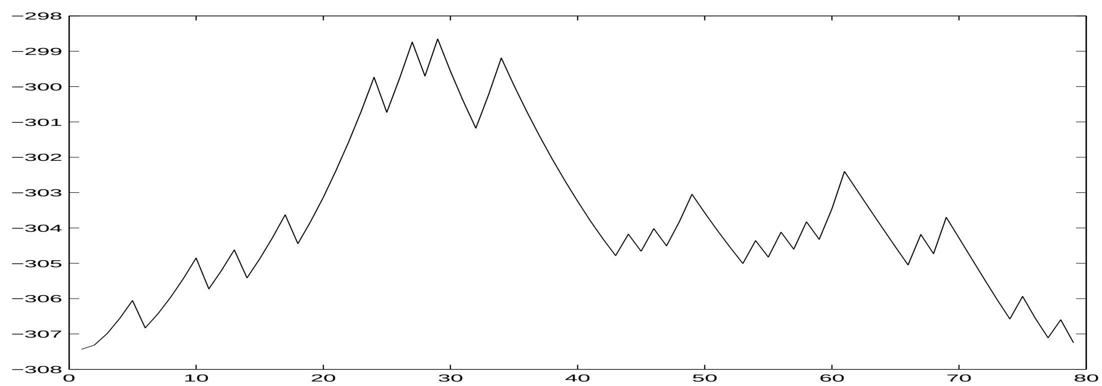
Figure 3.4: Changepoint procedure estimate



which implies that in the beginning, the hazard for patients receiving both therapies is about 4 times the hazard for those receiving only chemotherapy. However, for patients who survive after 262 days of chemotherapy and radiotherapy, their hazard is about half of the hazard for patients in the chemotherapy group.

One possible explanation of this phenomenon is that weaker patients receiving both chemotherapy and radiotherapy have a very high chance of death before the 262 days because of the strength of the treatment. However, those patients that are strong enough to survive the first period of treatment have a better chance of survival than patients receiving chemotherapy only.

Figure 3.5: MLE procedure estimate



Chapter 4

Discussion

4.1 Credit Risk Modeling and Analysis Using Copula Method

The 2007-2009 financial crisis drew a lot of attention to the securitized products of residential mortgages. There is plenty of literature analyzing the loss distribution of securitizations of mortgages (e.g., RMBS) such as the Pykhtin-Dev model, but there has not been much analyzing the loss distribution of re-securitizations (e.g., CDO of RMBS). Re-securitizations are treated more or less the same way as securitizations by the financial industry.

The first part of this thesis extended the Pykhtin-Dev model and derived analytically the loss distribution of re-securitizations under the ASRF framework, highlighting the important differences between re-securitizations and securitizations. Applying the model to credit rating of CDO tranches showed why it was unreasonable for the rating agencies to assign "AAA" ratings to CDOs made out of mezzanine tranches of subprime RMBS. We also proposed an alternative correlation structure for the securitization model, the Student's t Copula, instead of the commonly used Gaussian Copula. It was shown how tail correlation, being the main advantage of Student's t Copula, affects the loss distribution of securitization tranches, especially the tail loss.

For future work, it would be of interest to extend the Student's t Copula to re-securitzations. Computational efficiency will be key as the method tends to rely on Monte Carlo simulations.

4.2 Changepoint Approach to Survival Data

The lag effect and saturation effect problems are common and important problems in survival analysis. Previous studies of these problems focused on MLE based methods, such as Luo et al. (1994). They can be considered in fact to belong to a general class of changepoint problem where the treatment effect takes occasional jumps instead of staying constant over time. Changepoint problems have been studied extensively in statistical literatures, such as Lai and Xing (2011).

The second part of this thesis extended Lai and Xing's recent work in changepoint modeling, which was developed under a time series and Bayesian setup, to the lag effect problems in survival analysis. A general changepoint approach for Cox models was proposed. The procedure is flexible, allowing changepoint estimation with or without knowledge of the number of changepoints. Consistency of the estimator was proved under the fixed segmentation method. An example was given to illustrate how the procedure can be applied to clinical trial data to analyze time-varying treatment effect.

It would be valuable in future research to explore the asymptotic properties of the procedure under the more practical natural segmentation method. It would also be of interest to extend the approach possibly to regression models in survival analysis other than Cox model.

References

Altman, R. (2009). The great crash, 2008. *Foreign Affairs, 88*, 2-14.

Bai, J. (1997a). Estimation of a changepoint in multiple regression models. *Rev. Econ. Statist., 79*, 551-563.

Bai, J. (1997b). Estimation multiple breaks one at a time. *Econometric Theory, 13*, 315-352.

Bai, J., & Perron, P. (1998). Testing for and estimation of multiple structural changes. *Econometrica, 66*, 817-858.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics, 18*, 1-22.

Barry, D., & Hartigan, J. A. (1992). Product partition models for change point problems. *Ann. Statist., 20*, 260-279.

Barry, D., & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *J. Amer. Statist. Assoc., 88*, 309-319.

Basel Committee on Banking Supervision (2009). *Revisions to the Basel II market risk framework — final version.*

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist., 41*, 389-405.

Chernoff, H., & Zacks, S. (1964). Estimating the current mean of a Normal distribution which is subject to changes in time. *Ann. Statist., 35*, 999-1018.

Chib, S., Nardari, F., & Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *J. Econometrics, 108*, 281-316.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society Series B, 34*, 187-220.

Dev, A. (2004). *Economic Capital: A Practitioner Guide*. Risk Books.

Fang, H., & Fang, K. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal., 82*, 1-16.

Galiani, S. S. (2003). Copula functions and their application in pricing and risk managing multiname credit derivative products. *Master Thesis*, King's College London, UK.

Gordy, M. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation, 12*, 199-232.

Gorton, G. (2009). The panic of 2007, in *Maintaining Stability in a Changing Financial System*, Proceedings of the 2008 Jackson Hole Conference, Federal Reserve Bank of Kansas City (pp. 131-262).

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika, 82*, 711-732.

Hinkley, D. (1970). Inference about the change point in a sequence of random variables.

Biometrika, 57, 1-17.

Hull, J., & White, A. (2010). The risk of tranches created from residential mortgages. *Financial Analysts Journal, 66*(5), 54-67.

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*. Mcgraw-Hill.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association, 53*, 457-481.

Kotz, S., & Nadarajah, S. (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press.

Lai, T. L., & Xing., H. (2011). A simple Bayesian approach to multiple change-points.

Statistica Sinica, 21, 539-569.

Li, D. X. (2000). On default correlation: A Copula Function approach. *Journal of Fixed Income, 9*, 43-54.

Liu, J. S., & Lawrence, C. E. (1999). Bayesian inference on biopolymer models. *Bioinformatics, 15*, 38-52.

Luo, X. (1996). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics, 46*, 783-793.

Luo, X., Turnbull, B. W., Cai, H., & Clark, L. C. (1994). Regression for censored survival data with lag effects. *Communications in Statistics Ser. A., 23*(12), 3417-3438.

Luo, X. (1996). The asymptotic distribution of MLE of treatment lag threshold. *Journal of Statistical Planning and Inference, 53*(1), 33-61.

Luo, X., Turnbull, B. W., & Clark, L. C. (1997). Likelihood ratio tests for a changepoint with survival data. *Biometrika, 84*(3), 555-565.

Martinussen, T., & Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data.* Springer.

McCulloch, R. E., & Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Amer. Statist. Assoc., 88*, 968-978.

Miu, P., & Ozdemir, B. (2006). Basel requirements of downturn loss given default: modeling and estimating probability of default and loss given default correlations. *Journal of Credit Risk, 2*(2), 43-68.

Olshen, A. B., Venkatraman, E. S., Lucito, R., & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics, 5*, 557-572.

Pykhtin, M. (2004). Multi-factor adjustment. *Risk, 17*(3), 85-90.

Pykhtin, M., & Dev, A. (2002a). Analytical approach to credit risk modeling. *Risk, 15*(3), 26-32.

Pykhtin, M., & Dev, A. (2002b). Credit risk in asset securitizations: An analytical model. *Risk, 15*(5), 16-20.

Qu, Z., & Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica, 75*, 459-502.

Siegmund, D. (2004). Model selection in irregular problems: Applications to mapping quantitative trait loci. *Biometrika, 91*, 785-800.

Stablein, D. M., & Koutrouvelis, I. A. (1985). A two sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics, 41*, 643-652.

Vostrikova, L. J. (1981). Detecting disorder in multidimensional random processes. *Sov. Math. Doklady, 24*, 55-59.

Wang, J., & Zivot, E. (2000). A Bayesian time series model of multiple structural changes in level, trend, and variance. *J. Bus. Econ. Statist., 18*, 374-386.

Wilde, T. (2001). Probing granularity. *Risk, 14*(8), 102-106.

Yao, Y. (1988). Estimating the number of change-points via Schwarz criterion. *Statist. Probab. Lett., 6*, 181-189.

Zhang, N., & Siegmund, D. (2006). A modified Bayes information criterion with applications to comparative genomic hybridization data. *Biometrics, 63*, 22-32.

Zucker, D. M., & Lakatos, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika, 77*(4), 853-864.