TOKYO METROPOLITAN UNIVERSITY

Essays on Financial Markets: Pricing, Decision Making and Empirical Analysis

by

Tsz-Kin Chung

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Graduate School of Social Sciences
Department of Business Administration

August 2015

TOKYO METROPOLITAN UNIVERSITY

Abstract

Graduate School of Social Sciences
Department of Business Administration

Doctor of Philosophy

by Tsz-Kin Chung

This thesis is aimed to address several important issues in the financial market, including the decision making for trading, the pricing of financial instruments and the empirical analysis on bond yields. The topics are motivated by the new challenges in financial engineering as emerged from the recent financial crisis of 2008. In the first part, I study the optimal stopping problem for a short-selling strategy in an illiquid security lending market. In the second part, I study the equity-credit hybrid modeling and its application on the pricing of financial products including defaultable European options, capped variance swaps and contingent convertible bonds. Furthermore, I study the asymptotic approach to the multifactor stochastic volatility model and the pricing of index options in the S&P 500 and Nikkei 225 markets. In the third part, I estimate a non-linear term structure model that is able to capture the zero lower bound of interest rates.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor, Professor Keiichi Tanaka, for his patient guidance and continuous support during my doctoral study. I have learnt greatly from Professor Tanaka's mathematical rigor, succinct writing style and diligent approach to tackle research problems. His teaching and guidance have helped me greatly to understand many important concepts in probability theory and financial engineering.

I would like to thank Professor Hirokuni Iiboshi for his teaching on Bayesian econometrics as well as the joint project on estimating non-linear term structure modelling as an output. Professor Iiboshi is always kind to me. He is also extremely resourceful on econometrics techniques as well as frontier macro-economics problems.

I have benefited greatly from the stimulating academic environment and strong network of the TMU finance group. I would like to thank Professor Masaaki Kijima, Professor Yukio Muromachi and Professor Takeshi Shibata for various supports as well as the organization of conference and workshop. Professor Muromachi's class on risk management and Professor Shibata's class on contract theory have offered me an excellent opportunity to broaden my knowledge in financial engineering beyond my research area. Special thanks goes to Professor Kumiko Hattori at the Department of Mathematics and Information Sciences for offering me to join her classes on probability theory.

My sincere gratitude also goes to Professor Yue-Kuen Kwok at the Hong Kong University of Science and Technology for encouraging me to pursue PhD study in Japan and continue doing research on financial engineering. The chapters on equity-credit hybrid modeling and pricing of contingent convertibles are composed from the collaboration with Professor Kwok since I was studying the Master in Financial Mathematics in Hong Kong.

I am grateful to the organization of the RTG Summer School in Financial Mathematics at the Princeton University and the NUS-Santander Doctorate Workshop in Advanced Financial Risk Management at the National University of Singapore. I also thank the participants at the Quantitative Method in Finance (QMF) 2013, the 8-th Congress of Bachelier Society as well as the 2014 TMU Finance Workshop for the discussions and helpful comments.

I gratefully acknowledge the internship opportunities at the Hong Kong Institute for Monetary Research (HKIMR) and the Mitsubishi Trust Investment Technology Co. Ltd (MTEC), which offer me excellent environment to conduct research at the intersection in between academic and industry.

I would like to thank all of my friends and colleagues for their continuous support throughout my doctoral study in Japan.

The scholarship from the Asian Human Resources Fund granted by the Tokyo Metropolitan Government is gratefully acknowledged.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since the seminal paper by Black and Scholes (1973), the family of financial engineering models have been tested, challenged and improved throughout a number of crisis periods and the ever-changing macro-financial environment. The recent Lehman financial crisis in 2008 has highlighted the importance to model the market illiquidity, the interaction of equity risk and credit risk, as well as the term structure of interest rates close to the zero lower bound. Several important features of the Lehman crisis are the heightened counterparty risk exposure, evaporation of market liquidity and freezing of the funding and security lending market (for example, see Adrian et al., 2013; Brunnermeier, 2010; Duffie, 2010). The impairment of financial intermediations in turn led to abrupt fluctuations in asset prices, credit spreads and foreign exchange rates. Although it has been almost a decade since the outbreak of Lehman crisis, its impact on the financial market has been long lasting and has led to numerous reforms in the financial market structures and regulations. Furthermore, both the academic researchers and market participants have started looking for the next generation of financial engineering models that are capable to work well under the *new normal* going forward.

This thesis is aimed to address several important issues in the financial market, including the decision making for trading, the pricing of financial instruments and the empirical analysis on bond yields. The topics are motivated by the new challenges in financial engineering as emerged from the recent financial crisis of 2008. In the first part, I study the optimal stopping problem for a short-selling strategy in an illiquid security lending market. During a crisis period, a fund manager might find it very difficult to borrow a security for short-selling in order to hedge against the portfolio's market exposure. The introduction of the broker's recall is important when the depth and liquidity of the security lending market is thin and the continuation of a stock borrowing is not guaranteed. Moreover, as the borrowing cost increases due to the limited supply and

liquidity, it is necessary to take into account the loan fee which constitutes the running cost of holding a short position. These features make a short-selling very different from a buy-and-hold long position and we find that the investor's decision depends on the delicate balance between the benefit and the cost of holding the short position.

In the second part, I study the equity-credit hybrid modeling and its application on the pricing of financial products including defaultable European options, capped variance swaps and contingent convertible bonds. Empirical evidences show that the spillover and interaction in between equity risk and credit risk can be significant in particular during a crisis period. A joint equity-credit modeling is therefore crucial for the pricing of credit-sensitive equity derivatives and it allows one to develop a cross-asset class hedging strategy. I also extend the equity-credit hybrid modeling approach to the pricing of contingent convertible bond, which is a new financial instrument emerging after the Lehman crisis. Furthermore, I study the asymptotic expansion approach to multifactor Heston model and perform extensive calibration exercise to the S&P 500 and Nikkei 225 option markets in recent years. The mathematical models developed in this part are useful for the valuation and risk management of complex derivatives trading in the financial market.

In the third part, I estimate a non-linear term structure model when interest rates are close to the zero lower bound in the Japanese government bond market. The realistic modeling of the dynamics of bond yields under the zero interest rate policy is important for policy makers to gauge reliable information from the term structure of interest rates. In particular, I add in the macro-finance feature in order to explain the movements in bond yields with a richer economic interpretation and potentially improve the prediction of future bond yields by incorporating information beyond the bond market. The empirical study is useful for policy makers to better understand the market expectation and risk premium as embedded in the term structure of interest rates.

In the following, I review in more details the background and motivation for these topics.

1.1 Optimal Short-Covering of a Security

Background

The discussion on short-selling has been getting increasing attention among market participants, policy makers and academic researchers, in particular after the short-sale ban during the financial crisis in 2008 (Battalio and Schultz, 2011; Battalio et al., 2011). Short-selling is generally viewed as the market mechanism to correct the short-term deviation of security price from its fundamental value, and it plays an important role for

price discovery as well as market efficiency (Boehmer and Wu, 2013; Diether et al., 2009). In addition, short-selling has a wide range of applications in portfolio management, relative trading and risk management. Fund managers who want to tailor their market risk exposure often need to establish short positions to neutralize the portfolio's volatility, such as a market neutral strategy or a 130/30 long-short strategy. For relative trading strategies, such as pair trading and index arbitrage, investors have to take short positions in order to eliminate the unwanted common risk factors in the basket of stocks. For derivatives trading, option dealers need to delta-hedge the risk exposure of their option portfolio by maintaining dynamic short positions. In some applications for convertible securities, one can strip out the embedded options by shorting the underlying equity. It is estimated that short-selling trades constitute as much as 20% of New York Stock Exchange trading volume and almost 75% of the short-sellers are institutional investors (Diether et al., 2009).

In practice, a short-selling in the financial market can take two forms: the covered short-selling and the naked short-selling. In the covered short-selling, the short-seller has to borrow the security from a broker-dealer and sell it outright in the market. At some later time, the short-seller has to buy back the security and return it to the broker in order to close the position. On the other hand, a naked short-selling is often performed by market makers or high-frequency traders who square their positions in a short period, such as a daily basis (Battalio et al., 2011). As settlements usually take two to three business days, these investors can sell the security without actually borrowing it by only keeping the promise to deliver the security and square the position at the end of each trading day. The liquidity of the security lending market is of a less concern for these investors with a short-term horizon. In this thesis, we are interested in the optimal short-selling strategy implemented by an investor who has a relatively long investment horizon in terms of months or years. Moreover, the investor is intended to make a profit from a decline in the stock price, instead of taking short-term arbitrage opportunity. Therefore, we focus on the covered short-selling and assume that the investor must actually borrow the security from a broker-dealer before selling it in the market.

As mentioned, the study is motivated by the illiquidity and frictions in the security lending market during a crisis period. The impacts of short-selling constraints on stock and portfolio return has been well-documented in the empirical finance literature (Diether et al., 2009; Jones and Lamont, 2002). However, there has been no formal mathematical analysis on how various short-selling constraints might affect the decision making of a short-seller. In this thesis, we formulate the short-selling strategy as an optimal stopping problem and analysis how various short-selling constraints, including the recall risk from broker, the loan fee payment and the interest rate rebate, influence the decision making of a short-seller.

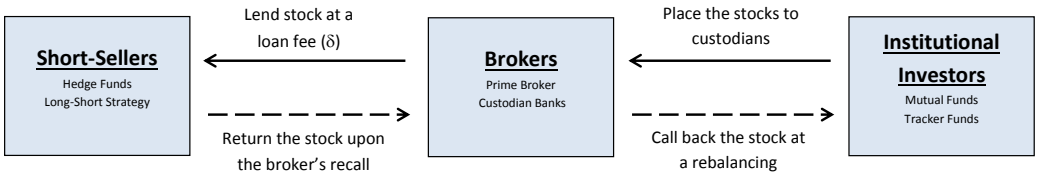


FIGURE 1.1: Mechanism of a short-selling transaction in the financial market.

Market mechanism and risk profile

The features in the security borrowing/lending market makes a short-selling very different from a buy-and-hold long position. Figure 1.1 illustrates the transaction and counterparties involved in a typical covered short-selling in the stock market. On the right hand side, we have some passive long-bias institutional investors, such as mutual funds, pension funds and tracker funds, who place the shares of stock into the broker-dealer who acts as the custodian for these investors. The broker-dealer, who holds the inventory of stocks, has the discretion to lend out the stock to other investors and earn the loan fee. On the left hand side, the short-seller, such as a hedge fund manager who implements a short-selling strategy, borrows the stock and sells it in the market. After that, the short-seller's objective is to "buy low, sell high" such that the stock is first sold high and purchased later at a lower price. The buying back and returning of the stock to the broker is called *short covering*. At the end, the short-seller makes a profit from a price decline or a loss from a price rise of the stock.

There are several real-world complications involved in the implementation of a short-selling strategy. A unique feature in a stock loan contract is that there is no guaranteed maturity and it is effectively rolled over on a daily basis as documented in D'Avolio (2002), Duffie et al. (2002) and Jones and Lamont (2002). Hence, the broker holds an option to recall the stock borrowing at any time. At a recall, the short-seller is then forced to cover the short position immediately, regardless of the mark-to-market profit or loss. This means the short-seller faces an uncertain (random) investment horizon in an illiquid security lending market in which replacing stocks are not immediately available. The risk of such an involuntary termination of a short-selling strategy is called the recall risk.

Besides the capital profit/loss associated with the recall risk, the short-seller also has to take into account the running cost and benefit of holding the short position. The broker charges the short-seller a loan fee, which is calculated as the loan fee rate times the stock price times the length of the period. It is worth to note that the loan fee is settled on a daily basis because both the short-seller and broker can effectively terminate the borrowing agreement at any time. At the same time, the short-seller deposits the sale

proceeds into a margin account, which generates interest income called the short interest rebate. When the interest rate is high, the income may become one of the return drivers of a short-selling strategy. Hence, the investor's decision depends on the balance between the benefit (interest income) and the cost (loan fee) of holding the short position.

1.2 Equity-Credit Hybrid Modeling

Background

The affine jump-diffusion (AJD) models have been widely used in the continuous time modeling of stochastic evolution of asset prices, bond yields and credit spreads. Some of the well known examples include the stochastic volatility (SV) model of Heston (1993), stochastic volatility jump-diffusion models (SVJ) of Bates (1996) and Bakshi et al. (1997), and stochastic volatility coherent jump model (SVCJ) of Duffie et al. (2000). The AJD models possess flexibility to capture the dynamics of market prices in various asset classes while admitting nice analytical tractability. For instance, the affine term structure models, which fall into the family of AJD models, have been frequently used to study the dynamics of bond yields and credit spreads (see Duffie and Singleton, 1999).

A number of studies have addressed the importance of including jump dynamics to valuation and hedging of derivatives. In the modeling of equity derivatives, Bakshi et al. (1997) illustrate that the stochastic volatility model augmented with the jump-diffusion feature produces a parsimonious fit to stock option prices for both short-term and long-term maturities. Empirical studies reported by Bates (1996), Pan (2002) and Erakar (2004) show that the inclusion of jumps in the modeling of stock price is necessary to reconcile the time series behavior of the underlying with the cross-sectional pattern of option prices. In particular, Erakar (2004) concludes from his empirical studies that simultaneous jumps in stock price and return variance are important in catering for different volatility regimes.

While the AJD models have been successfully applied in the valuation of both equity and credit derivatives, the joint modeling of equity and credit derivatives have not been fully addressed in the literature. Recently, a growing literature has highlighted such an interaction between equity risk (stock return and its variance) and credit risk (firm default risk). While the risk neutral distribution of stock return is fully conveyed by traded option prices of different strikes and maturities, the information of the arrival rate of default can be extracted from the bond yield spreads or credit default swap spreads. With the growing liquidity of the credit default swap (CDS) markets, the CDS spreads provide more reliable and updated information about the credit risk of

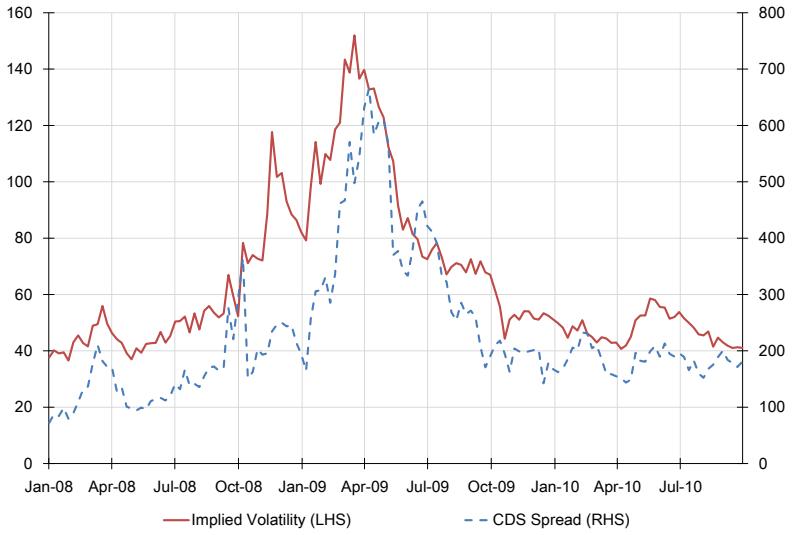


FIGURE 1.2: Implied volatility and CDS spread of Citibank around the Lehman crisis.

firms. Acharya and Johnson (2007) find that the CDS market contains forward looking information on equity return, in particular during times of negative credit outlooks. For equity options, Cremers et al. (2008), Zhang et al. (2009) and Cao et al. (2010) find that the out-of-the-money put options, which depict the negative tail of the underlying risk neutral distribution, are closely linked to yield spreads and CDS spreads of the reference firm. Figure 1.2 shows the implied volatility of 1-year at-the-money option (in %) and 5-year CDS spread (in bps) of the Citibank around the Lehman financial crisis. The strong empirical relationship motivates the joint modeling of equity and default risk, in particular during periods of financial distress.

Variance swaps with default risk

The pricing of exotic derivatives under the equity-credit model remains rare because of the loss in analytical tractability when path-dependent features are incorporated. Fortunately, it is recently found that the pricing of variance swaps and volatility derivatives can be made tractable by exploring the property of the characteristic function under the affine modeling framework. Variance swap is a financial instrument actively traded in the over-the-counter (OTC) market since last decade. At maturity, the buyer of a variance swap receives the difference between the realized variance over the contractual period and a fixed strike rate, which is called the variance swap rate. We refer the readers to Zheng and Kwok (2014) for an excellent review as well as the general pricing formula for a wide range of discretely-monitored exotic variance swaps, including the corridor variance swap and conditional variance swap.

The motivation of this work is to study the pricing of single-name variance swap in the presence of default risk. While previous studies have demonstrated the quantitative impact of stock price jumps (which are usually downside jumps) is significant to the pricing of variance swaps, the modeling of default risk has not been explicitly taken into account. It is well-known among market participants that single-name variance swap is exposed to default risk of the reference company and variance swap has been used as proxy instrument to hedge and trade credit risk - for example, to explore the arbitrage opportunity in between equity implied volatility versus credit spread written on the same company.[1] Having said that, before the outbreak of the financial crisis in 2008, most of the single-name variance swaps are referenced on high-grade companies in which the default probabilities are perceived to be small. Consequently, the impact of default risk has been largely ignored in the pre-crisis valuation of single-name variance swap.

Contingent convertibles

The contingent convertible bond (CoCo) is a hybrid fixed income security that provides a loss-absorption mechanism when the capital of the issuing bank falls close to the regulatory level as required by the Basel Committee on Banking Supervision. At a triggering event, the bond is automatically converted into the equity of the issuing bank (or an equivalent amount of cash). The conversion provides fresh capital to the issuing bank and saves it from financial distress. As a result, this can help to mitigate the chance of a systemic banking crisis while avoiding the use of taxpayer's money to bail out distressed financial institutions.

In principle, the pricing of CoCo is related to interest rate risk (due to the coupon and principal payments), equity risk (if the bond is converted to equity at the trigger event) and conversion/default risk. The conversion risk can be interpreted as the risk of an unfavorable conversion to a declined equity price that wipes off the value of the bond. The investment community typically considers the CoCo bond as a fixed income security as long as the tier-one capital ratio of the reference bank remains at a healthy level. When the capital ratio gets closer to the level of trigger such that the conversion probability is high, the risk profile of a CoCo bond should resemble a long equity position (similar to distressed fixed income security). Figure 1.3 shows that the CoCo price moves in tandem with the stock price (in GBP pence) during the sample period and this demonstrate the hybrid equity exposure of the CoCo bond. From a portfolio management perspective, it is important to better understand the decomposition of a CoCo bond into its fixed income and equity components.

[1] See: Variance swap: An introduction. Equity Derivative Strategy, Bear Stearn (2005) and Variance Swap, European Equity Derivatives Research, JP Morgan (2006).

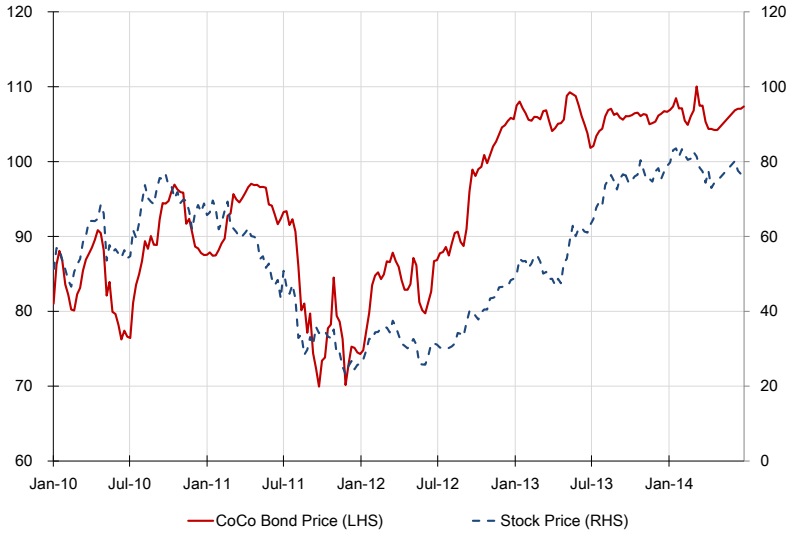


FIGURE 1.3: CoCo price and stock price of the Lloyds banking group since issuance.

1.3 Term Structure Modeling near the Zero Lower Bound

Background

Since the Lehman crisis in 2008, the US Federal Reserve has adopted various unconventional monetary policy tools as the nominal policy rate has attained its zero lower bound (ZLB). These policy tools include: (i) the large-scale asset purchases that aims to directly affect the interest rates by reducing the supply of treasury bonds and (ii) the forward guidance on the commitment of the duration of the zero interest rate policy that indirectly affects the interest rates through the influence on market expectation about the future path of the policy rate. A number of recent studies have examined the effectiveness of these policy tools by looking at the term structure of interest rates. For policy makers, it is crucial to extract the market expectation on future monetary policy using term structure models in two aspects:

- Gauging the likely path of future interest rate, i.e., forecasting based on the term structure model.
- Calculation of the term premium which allows one to disentangle from the yield curve the expected future interest rate and the risk premium as required by investors in the bond market.

In terms of modeling, the two aspects are linked to each other because one can calculate the term premium by subtracting the observed forward interest rate by the model-implied expected interest rate (i.e., the expected path of interest rate). Therefore, the

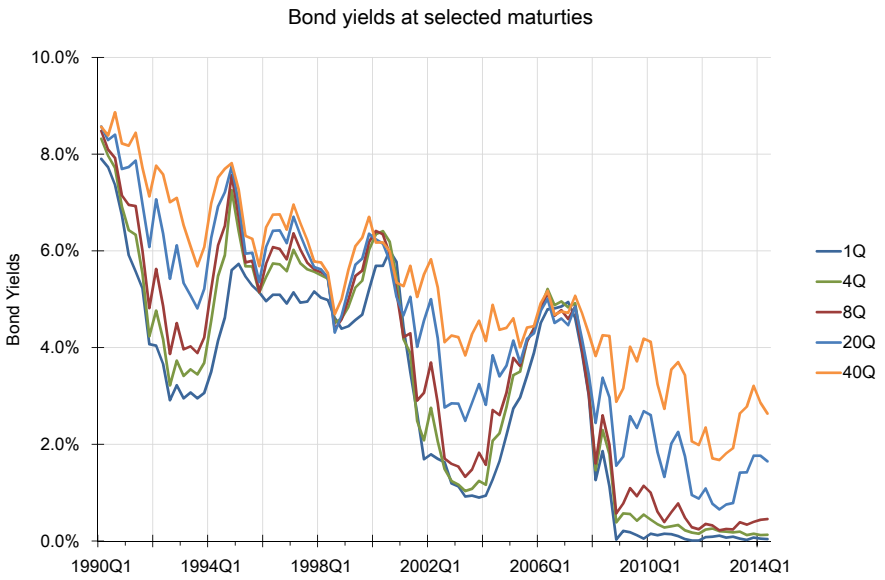


FIGURE 1.4: The US treasury bond yields including the Lehman financial crisis.

main concern is to compute the expected future interest rate under the physical probability measure instead of the risk-neutral measure as used for the no-arbitrage pricing of risk-free bonds. Such an exercise is essentially related to an accurate prediction of future bond yields and hence can be regarded as a forecasting problem. We demonstrate that a class of non-linear term structure model, known as the quadratic Gaussian term structure (QTSM) model, provides a highly flexible and analytical tractable framework to tackle this issue when interest rates are close to the ZLB. By using the log prediction score criteria as advocated by Geweke and Amisano (2011), we find that the QTSM provides a more realistic description of the dynamics of bond yields when interest rates are close to zero. The traditional Gaussian affine term structure model (ATSM) is inadequate to the modeling of interest rates near the ZLB. While we employ the Japanese government bond dataset in the estimation, our empirical results shed light on the future research on term structure modeling using the US treasury bond yield data.

In the literature, the modeling of the term structure of interest rates near the ZLB poses two major empirical challenges. Firstly, because nominal interest rates cannot fall below zero, the model has to preclude negative interest rates and generates an asymmetric distribution of interest rate near the ZLB. We note that the goodness-of-fit to the cross-sectional bond yields is not a guarantee of a good model because it may fail to capture the historical variations of bond yields in the time-series dimension. To this end, an appropriate model should also be able to capture the heteroscedasticity in the conditional volatility of bond yield, such as the compression of yield volatility as interest rates move towards the ZLB (Kim and Singleton, 2012). More importantly, the model should be capable to generate the stickiness feature of short rate to reflect the

persistence of zero interest rate monetary policy. For instance, the Japanese bond yields have been staying very close to zero for almost two decades.

ATSM

The ATSM has been a popular choice in the modeling of yield curve given its analytical tractable bond pricing formula as well as the linear dependence of the model-implied bond yields to the underlying factors or state variables (Piazzesi, 2010). The simple bond pricing formula and the linear structure of the model-implied bond yield makes the Gaussian ATSM easy to implement and estimate using various econometrics techniques. Nevertheless, both of the two aforementioned challenges cast doubt on the Gaussian ATSM as an appropriate model near the ZLB. In particular, the Gaussian distributional assumption of the ATSM near the ZLB suggests that its inferences on the expected path of future interest rate as well as the term premium can be heavily biased.

QTSM

An alternative is the QTSM as advocated by Ahn et al. (2002) and Leippold and Wu (2002), in which a lower bound of interest rates can be naturally imposed in the modeling framework. Indeed, the quadratic models have been widely adopted in the finance industry in the pricing and hedging of interest rate derivatives. It is important to note that the QTSM carries the same level of analytical tractability as the ATSM in which the bond pricing formula can be expressed in semi closed-form and involves only the calculation of a recursive relationship. Unfortunately, it appears that this class of model has been ignored by the academic literature and there is limited empirical study about it. Until recently, Kim and Singleton (2012) demonstrate in great details that the QTSM provides a good statistical description on Japanese bond yield under the countrys prolonged zero interest rate policy. Andreasen and Meldrum (2013) shows that the QTSM is able to capture the statistical feature of the US bond yields near the zero lower bound since 2008.

1.4 Mathematical Tools

In this thesis, I demonstrate the application of various tools in financial mathematics and financial econometrics, including optimal stopping theory, general affine process, credit risk modeling, term structure modeling and Bayesian estimation. As for mathematical technique, I make extensive use of stochastic calculus (Chapters 2-5) and specialized use of Malliavin calculus in asymptotic expansion (Chapter 6). Moreover, standard

mathematical analysis tools such as partial differential equations and ordinary differential equations are frequently employed during the solution procedure. For numerical techniques, I make use of Fourier transform analysis (Chapter 4), the integral equation approach (Chapter 5) and Markov Chain Monte Carlo technique (Chapter 7).

1.5 Copyrights

Several chapters in this thesis are based on published papers and I would like to acknowledge the permissions to allow inclusion of the following published papers in this thesis:

- Chapter 2 is published as [47]: Chung and Tanaka (2015) Optimal Timing for Short-Covering of an Illiquid Security. Journal of the Operations Research Society of Japan 58(2). The copyright of the paper is owned by the Operations Research Society of Japan.

- Chapter 3 is published as [45]: Chung (2015) Optimal Short-Covering with Regime Switching. To appear in Recent Advances in Financial Engineering 2014, edited by Masaaki Kijima, Yukio Muromachi and Takashi Shibata. The copyright of the paper is owned by the publisher World Scientific Publishing Co. Pte. Ltd.

- A small part (less than 20%) of the materials in Chapter 4 is published as [46]: Chung and Kwok (2014) Equity-Credit Modeling under Affine Jump-diffusion Models with Jump-to-Default. Journal of Financial Engineering 1(2). The copyright of the paper is owned by the publisher World Scientific Publishing Company.[2]

- Chapter 6 is published as [106]: Nagashima, Chung, and Tanaka (2014) Asymptotic Expansion Formula of Option Price under Multifactor Heston Model. Asia-Pacific Financial Markets 21(4):351–396. The copyright of the paper is owned by the publisher Springer Japan.

[2]Electronic version of an article published as [Journal of Financial Engineering, 1, 2, 2014, 25 pages][DOI: 10.1142/S2345768614500172][copyright World Scientific Publishing Company][http://www.worldscientific.com/worldscinet/jfe].

Part I

Decision Making

Chapter 2

Optimal Timing for Short-Covering of an Illiquid Security

2.1 Introduction

Short-selling is the selling of a financial security that the investor does not own. The trading provides an efficient means for investors to exploit the opportunity or hedge against downside risk when they anticipate the overpricing of a security and speculate its future decline in value. A typical situation involving short-selling transactions can be described as follows. Some institutional investors are long biased and hence only rebalance their portfolios on a quarterly or yearly basis. These institutional investors are usually mutual funds, pension funds or tracker funds. They place their stocks with the broker who acts as a custodian. The broker who holds the inventory of stocks has the discretion to lend out the stock in order to earn a loan fee income. On the other side is the short-seller (e.g., a hedge fund manager) who implements a short-selling position by borrowing the stock ("stock loan") from such a broker and selling it in the market. After that, the short-seller's objective is to "buy low, sell high" such that the stock is first sold high and purchased later at a lower price. The buying back and returning of the stock to the broker is called *short covering*. At the end, the short-seller makes a profit from a price decline or a loss from a price rise of the stock.

There are a number of real-world complications involved in the implementation of a short-selling strategy. A unique feature in a stock loan contract is that there is no guaranteed maturity and it is effectively rolled over on a daily basis as documented in D'Avolio (2002). Hence, the broker holds an option to recall the stock borrowing at any

time. At a recall, the short-seller is then forced to cover the short position immediately, regardless of a profit or loss, if replacing stocks are not found and placed with the broker. The risk of such an involuntary termination of a short-selling strategy is called the recall risk. Besides the capital profit/loss associated with the recall risk, the short-seller also has to take into account the running cost of the strategy. The broker charges the short-seller a loan fee, which is calculated as the loan fee rate times the stock price times the length of the period. As noted in D'Avolio (2002), the loan fee rate varies dramatically across different categories of stocks from 50 to 800 basis points. At the same time, the short-seller deposits the sale proceeds into a margin account, which generates interest income called the short interest rebate. When the interest rate is high, the income may become one of the return drivers of a short-selling strategy. Hence, the investor's decision depends on the balance between the benefit (interest income) and the cost (loan fee) of holding the short position.

The optimal trading rule of a long position has been formulated as an optimal stopping problem in Peskir and Shiryaev (2005) and extended in Guo and Zhang (2005), in which the investor initially holds a security and seeks the optimal timing to sell it in order to maximize the expected discounted payoff. In this paper, we formulate a short-selling strategy as an optimal stopping problem and seek the optimal timing of the short covering in the presence of a recall risk, loan fee and interest income. The feature of the random recall gives rise to an optimal stopping problem with a random time horizon. We apply the resolvent operator to simplify the problem and derive the solution based on the approach in Dayanik and Karatzas (2003). As such, we directly construct the solution rather than guessing the form of the value function and stopping rule. One of the interesting results is that, depending on the levels of loan fee and interest rates, the optimal stopping problem is either of a put-type problem with a down-and-out stopping rule or a call-type problem with an up-and-out stopping rule. We find that the recall risk has a significant impact on the value function and the corresponding optimal threshold. The value function may become negative because of the possibility of a forced termination, and the short-seller is likely to stop earlier at the closer optimal threshold to the entry price as a result of the random recall (the put-type problem) or the relatively expensive net running cost of keeping the position (the call-type problem). Given the closed-form solution, we characterize explicitly the investor's active region in terms of the loan fee rate and interest rate in several ways. We show that the active region depends sensibly on the stock price volatility, expected return and recall intensity.

2.2 Model Setup

We fix the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and assume a stock price to be a one-dimensional diffusion process $X = (X_t)_{t\geq 0}$ satisfying

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \qquad X_0 = x > 0. \tag{2.1}$$

Here, $(W_t)_{t\geq 0}$ is a standard Brownian motion, μ is the expected return of the stock and σ is the volatility of the stock. The infinitesimal generator of the stock price process X is given by

$$\mathcal{L}_X = \frac{1}{2}\sigma^2 x^2 \frac{d^2}{dx^2} + \mu x \frac{d}{dx}.$$

2.2.1 Short-Seller's Problem

At time $t = 0$, an investor makes or keeps a short position of the stock whose sale proceeds is K. For such a short position, we assume that there is a stock loan (or securities lending) market to borrow/lend the stock against the loan fee, although the liquidity may be limited[1] in the sense that the loan contract is available only with a specific broker because of the illiquidity of the stock loan market. The contract may be automatically renewed instantaneously, although there is a chance that the broker will not be able to find stock to replace. Hence, the lender does not renew the contract at the broker's recall time τ_R which is an exponential random variable $Exp(\lambda)$ with parameter $\lambda \geq 0$ independent of the stock price process.[2] Once the loan contract is terminated, the short-seller has to cover the short position by buying stock at the market price.

We write $\mathcal{F}_t^W = \sigma(W_s; s \in [0,t])$, $\mathbb{F}^W = \left(\mathcal{F}_t^W\right)_{t\geq 0}$ and $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, and assume that $\mathcal{F}_t^W \vee \sigma(1_{\{\tau_R > t\}}) \subset \mathcal{F}_t$. We denote the expectation $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot \,|\, X_0 = x,\, \tau_R > 0]$ under $\mathbb{P}$.

The loan fee is charged instantaneously based on the current stock price, i.e., the borrower makes the loan fee payment $\delta X_t dt$ over a small time interval dt, where δ is the constant loan fee rate. The short-seller deposits the initial proceeds K from selling the stock into a margin account that pays interest continuously at a constant rate q. As a result, the net cash outflow is given by $(\delta X_t - qK)\, dt$ over a time interval dt, which can be positive or negative depending on the levels of the stock price, the loan fee rate and

[1] A repo contract is similar to a stock loan contract. They differ in terms of cash flow. In a stock loan contract only the interest equivalent cash flow (loan fee) is paid to the lender instantaneously without any payment of the notional amount.

[2] The exponential variable is a popular choice for the modeling of random arrival times in finance and economics (see Guo and Liu, 2005; Pliska and Ye, 2007).

the interest rate. The net cash flow $\delta X_t - qK$ is sometimes referred to as the effective loan fee and can be interpreted as the net running cost of the short-selling strategy.

The short position of the investor is kept until she buys back at the market price either at her own discretion or following a recall by the broker. K may or may not be equal to $X_0 = x$. She seeks the optimal timing of short covering at her own discretion. The short-seller's problem is to optimize the expected net profit discounted at her own discount rate $\beta > \max(\mu, 0)$

$$v(x) = \sup_{\tau \in \mathcal{A}} \mathbb{E}_x \left[e^{-\beta(\tau \wedge \tau_R)} \left(K - X_{\tau \wedge \tau_R} \right) - \int_0^{\tau \wedge \tau_R} e^{-\beta s} \left(\delta X_s - qK \right) ds \right], \tag{2.2}$$

where $\mathcal{A}$ is the set of all $\mathbb{F}^W$-stopping times taking values in $[0, \infty]$. Note that a higher (lower) discount rate β indicates that an investor is less (more) patient.

For a strong Markov process Y, we denote the resolvent operator by

$$\left(\mathcal{R}_\beta^Y f \right)(x) = \mathbb{E} \left[\int_0^\infty e^{-\beta s} f\left(Y_s\right) ds \middle| Y_0 = x \right],$$

which is useful for the evaluation of the total cost of keeping the short-position (see Rogers and Williams, 2000).

Lemma 2.1. *(1) For any stopping time τ, it holds that*

$$\mathbb{E} \left[\int_\tau^\infty e^{-\beta s} f\left(Y_s\right) ds \middle| Y_0 = x \right] = \mathbb{E} \left[e^{-\beta \tau} \left(\mathcal{R}_\beta^Y f \right)\left(Y_\tau\right) \middle| Y_0 = x \right].$$

(2) Suppose that an exponentially distributed random variable U with parameter λ is independent of Y. For any stopping time τ, it holds that

$$\mathbb{E} \left[e^{-\beta(\tau \wedge U)} f\left(Y_{\tau \wedge U}\right) \middle| Y_0 = x \right] = \mathbb{E} \left[e^{-(\beta+\lambda)\tau} \left(f - \phi \right)\left(Y_\tau\right) \middle| Y_0 = x \right] + \phi\left(x\right),$$

where

$$\phi\left(x\right) = \mathbb{E}_x \left[e^{-\beta U} f\left(Y_U\right) \middle| Y_0 = x \right].$$

(3) Suppose that $dY_t = \mu Y_t dt + \sigma Y_t dW_t$ and $f\left(x\right) = kx + l$. When $\beta > \mu$,it holds that

$$\left(\mathcal{R}_\beta^Y f \right)(x) = \frac{k}{\beta - \mu} x + \frac{l}{\beta}.$$

Proof. See Appendix A. □

By Lemma 2.1, we can re-write the short-seller's problem as

$$v(x) = \sup_{\tau \in \mathcal{A}} \mathbb{E}_x \left[e^{-(\beta+\lambda)\tau} \left(g - \phi \right) \left(X_\tau \right) \right] - \left(\mathcal{R}_\beta^X f \right)(x) + \phi(x), \tag{2.3}$$

where

$$f(x) = \delta x - qK, \quad g(x) = K - x + \left(\mathcal{R}_\beta^X f \right)(x), \quad \phi(x) = \lambda \left(\mathcal{R}_{\beta+\lambda}^X g \right)(x).$$

See Appendix B for the derivation, where we used $\beta > \mu$ and $\beta + \lambda > \mu$ to calculate the resolvent operators acting on f and g respectively. Here,

$$u(x) = \sup_{\tau \in \mathcal{A}} \mathbb{E}_x \left[e^{-(\beta+\lambda)\tau} \left(g - \phi \right)(x) \right], \tag{2.4}$$

is an infinite horizon problem with the linear payoff function

$$\left(g - \phi \right)(x) = \frac{\beta}{\beta + \lambda} \left(1 - \frac{q}{\beta} \right) K - \frac{\beta - \mu}{\beta + \lambda - \mu} \left(1 - \frac{\delta}{\beta - \mu} \right) x.$$

The coefficients $\rho = 1 - \dfrac{\delta}{\beta - \mu}$ and $\eta = 1 - \dfrac{q}{\beta}$ play important roles in our analysis. A sufficiently low loan fee rate implies that the gain function $g - \phi$ is a decreasing function of x, while in an expensive loan fee environment it is an increasing function. Hence, it is not so straight-forward to reduce the problem to a perpetual American call or put problem. Apparently, the investor's optimal strategy depends on the magnitudes of the loan fee rate and the interest rate via the signs of ρ and η, which lead to different signs of the slope and intersection of the payoff function. A patient investor's η tends to be negative, while an impatient one's η tends to be positive. $\eta \gtrless 0$ is equivalent to $q \lessgtr \beta$ and $\rho \gtrless 0$ is equivalent to $\delta \lessgtr \beta - \mu$.

2.2.2 Solution

We apply the results of Dayanik and Karatzas (2003) to obtain the solution of the auxiliary problem (2.4). For the readers' convenience, we briefly review the procedure in Dayanik and Karatzas (2003) to solve an optimal stopping problem

$$V(x) = \sup_{\tau \in \mathcal{A}} \mathbb{E}_x \left[e^{-\alpha\tau} h\left(X_\tau \right) \right], \tag{2.5}$$

for a geometric Brownian motion X given by (2.1), the linear payoff $h(x) = kx + l$ and $\alpha > \mu$. Let ψ and φ be the solutions of $\left(\mathcal{L}_X - \alpha \right) u = 0$ on $(0, +\infty)$ such that (i) ψ is a positive increasing function and (ii) φ is a positive decreasing function. Namely, they

are

$$\psi(x) = x^n, \quad \varphi(x) = x^m,$$

where $n > 1$ and $m < 0$ are the distinct roots of

$$p(x) = \frac{1}{2}\sigma^2 x(x-1) + \mu x - \alpha$$

since $p(-\infty) > 0$, $p(+\infty) > 0$, $p(0) < 0$ and $p(1) < 0$. Then, we take

$$F(x) = \frac{\psi}{\varphi}(x) = x^\theta, \quad F^{-1}(y) = y^{1/\theta}, \quad \theta = n - m > 1,$$

such that F is positive and increasing. It can be checked that

$$l_0 = \limsup_{x \downarrow 0} \frac{\max(h(x), 0)}{\varphi(x)} = 0, \quad l_\infty = \limsup_{x \uparrow +\infty} \frac{\max(h(x), 0)}{\psi(x)} = 0.$$

Define the function $H : [0, +\infty) \to \mathbb{R}$ as

$$H(y) = \left\{ \begin{array}{ll} \dfrac{h\left(F^{-1}(y)\right)}{\varphi\left(F^{-1}(y)\right)}, & \text{when } y > 0 \\ l_0, & \text{when } y = 0 \end{array} \right\} = y^{-m/\theta}\left(ky^{1/\theta} + l\right).$$

We denote by $W : [0, +\infty) \to \mathbb{R}$ the smallest non-negative concave majorant of H. The candidate stopping time is the first-passage-time defined as

$$\tau^* = \inf\left\{t \geq 0; X_t \in \Gamma\right\}, \tag{2.6}$$

where $\Gamma = \{x \in \mathbb{R}_+ : V(x) = h(x)\}$ is the stopping region. The solution to the optimal stopping problem (2.5) can be constructed as follows.

Lemma 2.2 (Dayanik and Karatzas (2003): Prop. 5.10, 5.12, 5.13)**.** *(1) If l_0, l_∞ are finite, the value function V is finite and continuous on $\mathbb{R}_+$ and it is given by*

$$V(x) = \varphi(x) W\left(F(x)\right).$$

(2) If h is continuous and $l_0 = l_\infty = 0$, the stopping time τ^* of (2.6) is the optimal stopping time.

The key of the result is the construction of the smallest non-negative concave majorant W of H. To this end, we need to investigate the property of H on $[0, +\infty)$. When $l/k < 0$, we see that

$$H(0) = H(y_0) = 0, \quad H'(y_1) = 0, \quad H''(y_2) = 0,$$

where

$$y_0 = \left(-\frac{l}{k}\right)^\theta, \quad y_1 = \left(-\frac{l}{k}\frac{m}{m-1}\right)^\theta, \quad y_2 = \left(-\frac{l}{k}\frac{n}{n-1}\frac{m}{m-1}\right)^\theta.$$

Note that $y_1 < y_0$ and $y_1 < y_2$. For different values of l and k, the analysis can be split into four cases:

1. When $k < 0$ and $l > 0$, we have

   - $H(0) = 0$, $H(+\infty) = -\infty$ and $H'(y) \to +\infty$ as $y \downarrow 0$.
   - $H(y) > 0$ for $0 < y < y_0$ and $H(y) < 0$ for $y > y_0$.
   - H is increasing concave on $(0, y_1)$, decreasing concave on (y_1, y_2) and decreasing convex on $(y_2, +\infty)$.

   Hence, W can be constructed as

   $$W(y) = \begin{cases} H(y), & 0 \le y \le y_1, \\ H(y_1), & y_1 < y. \end{cases}$$

   This corresponds to a put-type problem as shown in Figure 1(a).

2. When $k > 0$ and $l < 0$, we see

   - $H(0) = 0$, $H(+\infty) = +\infty$ and $H'(y) \to -\infty$ as $y \downarrow 0$.
   - $H(y) < 0$ for $0 < y < y_0$ and $H(y) > 0$ for $y > y_0$.
   - H is decreasing convex on $(0, y_1)$, increasing convex on (y_1, y_2) and increasing concave on $(y_2, +\infty)$.

   Hence, W is given by

   $$W(y) = \begin{cases} \dfrac{H(y^*)}{y^*} y, & 0 \le y < y^*, \\ H(y), & y^* \le y, \end{cases}$$

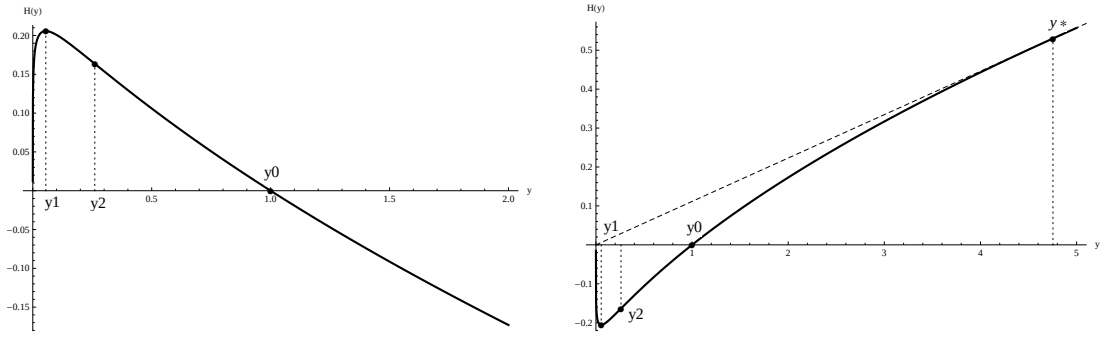
   where $y^* > y_2$ is determined by a tangent line of H from the origin such that $H(y^*) = y^* H'(y^*)$, that is

   $$y^* = \left(-\frac{l}{k}\frac{n}{n-1}\right)^\theta.$$

   This corresponds to a call-type problem as shown in Figure 1(b).

3. When $k \ge 0$ and $l \ge 0$ but $(k, l) \ne (0, 0)$, we see that H is increasing concave on $(0, +\infty)$. Hence, we take $W(y) = H(y)$.

4. When $k < 0$ and $l < 0$, we have $h(x) < 0$ for all x and it is trivial to take $\tau = \infty$ (see Section 3 Dayanik and Karatzas, 2003).



(A) When $k < 0$ and $l > 0$. (B) When $k > 0$ and $l < 0$.

FIGURE 2.1: Graph of H for different values of k and l.

By applying Lemma 2.2, we obtain the following results.

Proposition 2.3. *The auxiliary optimal stopping problem (2.4) has the solution:*

Case (a) *When* $\rho \geq 0$, $\eta \leq 0$, $u = 0$, $\tau^* = \infty$,

Case (b) *When* $\rho > 0$, $\eta > 0$, $u = u_p$, $\tau^* = \inf\{t \geq 0; X_t \leq b_p\}$, $b_p = \dfrac{m}{m-1} aK$,

Case (c) *When* $\rho \leq 0$, $\eta \geq 0$, $u = g - \phi$, $\tau^* = 0$,

Case (d) *When* $\rho < 0$, $\eta < 0$, $u = u_c$, $\tau^* = \inf\{t \geq 0; X_t \geq b_c\}$, $b_c = \dfrac{n}{n-1} aK$,

where

$$a = \frac{\beta}{\beta + \lambda} \frac{\beta + \lambda - \mu}{\beta - \mu} \frac{\eta}{\rho}, \tag{2.7}$$

satisfying $(g - \phi)(aK) = 0$, *and*

$$u_p(x) \triangleq \begin{cases} (g-\phi)(b_p) \left(\dfrac{x}{b_p}\right)^m, & x \in (b_p, \infty), \\ (g-\phi)(x), & x \in (0, b_p], \end{cases} \tag{2.8}$$

$$u_c(x) \triangleq \begin{cases} (g-\phi)(x), & x \in [b_c, \infty), \\ (g-\phi)(b_c) \left(\dfrac{x}{b_c}\right)^n, & x \in (0, b_c). \end{cases} \tag{2.9}$$

Proof. See Appendix C. □

By rearranging (2.3) and (2.4), we obtain the following main result.

Theorem 2.4. *The value function v of the optimal stopping problem (2.2) is given by*

$$v(x) = u(x) + \phi(x) - g(x) + K - x, \quad x \in \mathbb{R}_+.$$

2.2.3 Discussion

From Proposition 2.3 and Theorem 2.4, $v(x)$ has the two lower bounds: $h_0(x) = K - x$ corresponding to $\tau^* = 0$ (immediate exercise) and $h_\infty(x) = \phi(x) - g(x) + K - x$ corresponding to $\tau^* = \infty$ (wait and see). Figure 2.2 illustrates what is happening in each case.

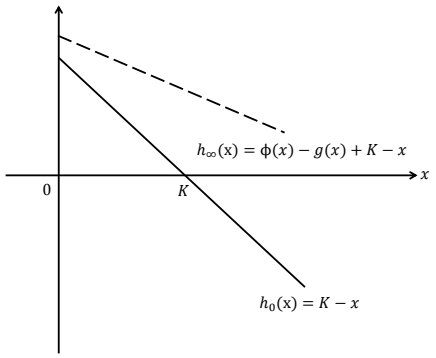
- Case (a): $h_\infty(x)$ is always greater than $h_0(x)$ thanks to the cheap loan fee and the high interest income. It is optimal to wait and see until the recall time ($\tau^* = \infty$).
- Case (b): The net running cost of the short position is relatively cheap. It is better to stop than wait when x is sufficiently small enough to satisfy at least $h_\infty(x) < h_0(x)$. The optimal strategy is to cover the short position at the threshold b_p, i.e., $\tau^* = \inf\{t \geq 0; X_t \leq b_p\}$.
- Case (c): $h_0(x)$ is always greater than $h_\infty(x)$ because of the expensive loan fee and the low interest income. It is optimal to stop immediately and hence no entry occurs ($\tau^* = 0$).
- Case (d): The net running cost of the short position is relatively expensive. It is better to stop than wait when x is sufficiently large enough to satisfy at least $h_\infty(x) < h_0(x)$. The optimal strategy is to cover the short position at the threshold b_c, i.e., $\tau^* = \inf\{t \geq 0; X_t \geq b_c\}$.

In the put-type problem, the short covering at the threshold b_p involves taking profit while the short covering at b_c in the call-type problem involves a loss cut in the capital[3]. In either case, net interest income until the short covering depends on the historical path. The mandatory short covering upon recall yields another opportunity of profit taking or loss cut. Compared with a case without a random recall risk, the short-seller is likely to stop earlier at an optimal threshold closer to the entry price because of the random recall (an early profit taking on the put-type problem) or the relatively expensive net running cost of keeping the position (an early loss cut on the call-type problem). Note that in Cases (b) and (d), as illustrated in Figure 2.2, it holds that

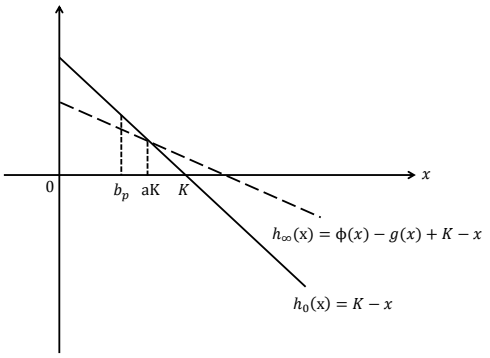
$$\text{Case (b)} \qquad b_p < aK, \qquad g(b_p) - \phi(b_p) = \frac{1}{1-m}\frac{\beta}{\beta+\lambda}\eta K > 0,$$

$$\text{Case (d)} \qquad aK < b_c, \qquad g(b_c) - \phi(b_c) = \frac{1}{1-n}\frac{\beta}{\beta+\lambda}\eta K > 0,$$

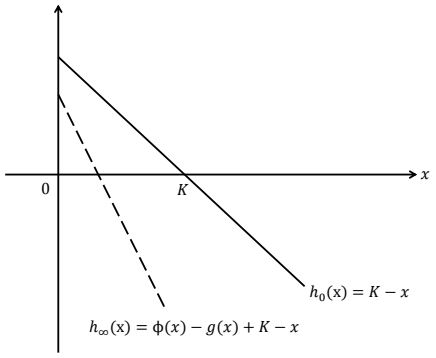
[3]This is a voluntary loss cut in order to avoid the high running cost as the stock price increases.



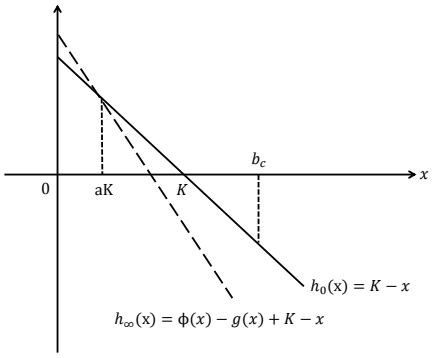
(A) When $\rho \geq 0$ and $\eta \leq 0$, the optimal strategy is $\tau^* = \infty$.



(B) When $\rho > 0$ and $\eta > 0$, the optimal strategy is $\tau^* = \inf\{t \geq 0; X_t \leq b_p\}$.



(C) When $\rho \leq 0$ and $\eta \geq 0$, the optimal strategy is $\tau^* = 0$.



(D) When $\rho < 0$ and $\eta < 0$, the optimal strategy is $\tau^* = \inf\{t \geq 0; X_t \geq b_c\}$.

FIGURE 2.2: The lower bounds h_0 and h_∞ of the value function v.

because $m < 0$ and $n > 1$. It can be seen that

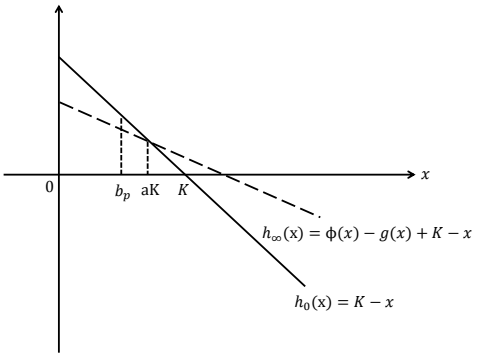
$$\text{Case (b)} \qquad u_p(x) = (g(b_p) - \phi(b_p)) \left(\frac{x}{b_p}\right)^m > 0, \quad x \in (b_p, \infty),$$

$$\text{Case (d)} \qquad u_c(x) = (g(b_c) - \phi(b_c)) \left(\frac{x}{b_c}\right)^n > 0, \quad x \in (0, b_c),$$

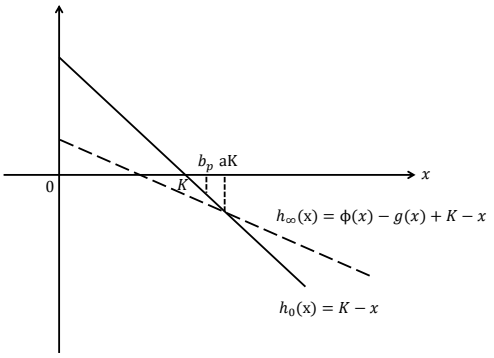
represent the value of the short-seller's optionality to stop at a finite time. The quantities $(x/b_p)^m$ and $(x/b_c)^n$ are the expected present values of one dollar when the stock price hits the thresholds b_p and b_c, respectively, before the random recall. The corresponding dollar amounts at the exercises are $g(b_p) - \phi(b_p)$ and $g(b_c) - \phi(b_c)$, respectively.

2.3 Active Condition

We need to consider the condition that the short-seller actually enters into a short position. As Case (a) and Case (c) are trivial, let us restrict our attention to the put-type problem (Case (b)) and the call-type problem (Case (d)). In order that the investor



(A) When $\rho > 0$, $\eta > 0$ and $a \leq 1$, then $b_p < aK \leq K$ always holds.



(B) When $\rho > 0$, $\eta > 0$ and $a > 1$. If $a < (m-1)/m$ then $b_p < K$ while $K \leq b_p$ in the case of $a \geq (m-1)/m$.

FIGURE 2.3: The active condition and optimal threshold for the put-type problem.

is active in the sense of waiting for the stock price's passage over the thresholds obtained in the previous subsection until the random recall, it is clear that $K > b_p$ on the put-type problem and $K < b_c$ on the call-type problem must hold. They are rewritten in terms of a, m and n in the next proposition.

Proposition 2.5. *For the put-type problem ($\rho > 0$, $\eta > 0$), the active condition $K > b_p$ is equivalent to the condition $a < (m-1)/m$. For the call-type problem ($\rho < 0$, $\eta < 0$), the active condition $K < b_c$ is equivalent to the condition $a > (n-1)/n$.*

Figure 2.3 illustrates the active condition for the put-type problem. When $a < 1$, we see $b_p < aK < K$ and the active condition $K > b_p$ is always satisfied. When $a > 1$, however, the active condition $K > b_p$ is satisfied only when $a < (m-1)/m$. Clearly, the active condition depends on the coefficients η, ρ and other model parameters. The active region on the (η, ρ)-plane can be characterized explicitly as

$$\rho > \begin{cases} \dfrac{m}{m-1}\gamma\eta, & \eta \geq 0, \\ \dfrac{n}{n-1}\gamma\eta, & \eta < 0, \end{cases} \qquad \text{where } \gamma = \frac{\beta}{\beta+\lambda}\frac{\beta+\lambda-\mu}{\beta-\mu},$$

as shown in Figure 2.4(a). They can be converted on the (q, δ)-plane as

$$\delta - (\beta - \mu) < \begin{cases} \dfrac{n}{n-1}\dfrac{\beta+\lambda-\mu}{\beta+\lambda}(q-\beta), & q > \beta, \\ \dfrac{m}{m-1}\dfrac{\beta+\lambda-\mu}{\beta+\lambda}(q-\beta), & q \leq \beta, \end{cases}$$

as in Figure 2.4(b). On the (q, δ)-plane, the boundaries of the two conditions in Proposition 2.5 are graphically represented by the two straight lines L_1 and L_2 in Figure 2.4(b). The slopes M and N of L_1 and L_2, respectively, are represented solely by the roots n and m, respectively, as in the next proposition, which also shows that a change of the

model parameters leads to an expansion or a contraction of the active region as the two lines rotate around $(\beta, \beta - \mu)$.

Proposition 2.6. *(1) The slopes on the (q, δ)-plane*

$$M = \frac{m}{m-1}\frac{\beta+\lambda-\mu}{\beta+\lambda}, \quad N = \frac{n}{n-1}\frac{\beta+\lambda-\mu}{\beta+\lambda}$$

have the following properties

$$N = \frac{m-1}{m}, \quad M = \frac{n-1}{n},$$

and

$$\frac{\partial N}{\partial \sigma^2} > 0, \quad \frac{\partial N}{\partial \mu} < 0, \quad \frac{\partial N}{\partial \lambda} < 0, \qquad \frac{\partial M}{\partial \sigma^2} < 0, \quad \frac{\partial M}{\partial \mu} < 0, \quad \frac{\partial M}{\partial \lambda} > 0.$$

(2) The thresholds b_p, b_c have the following properties

$$\frac{\partial b_p}{\partial \lambda} > 0, \qquad \frac{\partial b_c}{\partial \lambda} < 0.$$

Proof. n, m are solutions of the quadratic equation $(1/2)\sigma^2 x(x-1) + \mu x - (\beta+\lambda) = 0$. Then it holds that

$$\frac{\beta+\lambda-\mu}{\beta+\lambda} = \frac{(n-1)(m-1)}{nm},$$

and

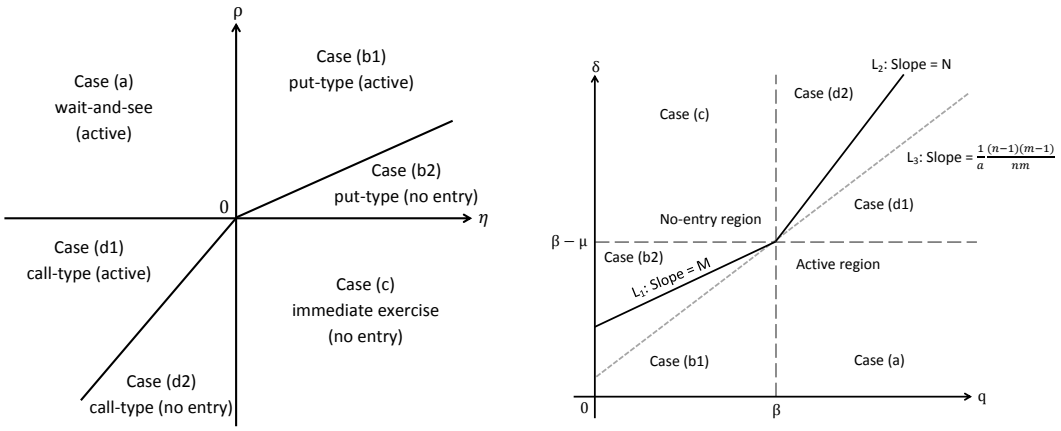
$$\frac{\partial m}{\partial \sigma^2} > 0, \qquad \frac{\partial m}{\partial \mu} < 0, \qquad \frac{\partial m}{\partial \lambda} < 0, \qquad \frac{\partial n}{\partial \sigma^2} < 0, \qquad \frac{\partial n}{\partial \mu} < 0, \qquad \frac{\partial n}{\partial \lambda} > 0.$$

□

Proposition 2.6 (1) implies that the active region expands as the stock volatility σ increases (more opportunistic trade) and contracts as the recall intensity λ increases (higher recall risk) when the loan fee rate δ and the interest rate q are fixed. The separating lines rotate clockwise when μ increases. With an increase of λ, the thresholds become closer to the original sales price so that the short-seller will close the position more quickly and more conservatively.

Finally, to represent the active condition graphically in relation to L_1 and L_2, we rewrite the definition of a in (2.7) as

$$\delta - (\beta - \mu) = \frac{1}{a}\frac{(n-1)(m-1)}{nm}(q - \beta),$$



(A) Active region on the (η, ρ)-plane. (B) Active region on the (q, δ)-plane.

FIGURE 2.4: Investor's active region. The origin on the (η, ρ)-plane corresponds to the point $(\beta, \beta - \mu)$ on the (q, δ)-plane.

which is drawn as the dashed line L_3 in Figure 2.4(b). In Case (b) on the domain $q \leq \beta$ in the graph, the active condition $a < (m-1)/m$ is equivalent to the condition

$$\frac{1}{a}\frac{(n-1)(m-1)}{nm} > \frac{n-1}{n}, \tag{2.10}$$

which implies that the slope M (the RHS of (2.10)) of line L_1 must be smaller than the slope of L_3 (the LHS of (2.10)). Similarly, for Case (d), $a > (n-1)/n$ is equivalent to the condition that the slope N of line L_2 is larger than the slope of L_3. In other words, the active condition is equivalent to the condition that the line L_3 lies below the lines L_1 and L_2.

2.4 Numerical Examples

In this section, we conduct a comparative analysis so as to understand the properties of the decision making using numerical examples for various levels of recall intensity, loan fee rate and interest rate. Unless otherwise stated, we assume the following parameter values: the short-seller's discount rate $\beta = 0.05$, the stock price volatility $\sigma = 0.2$, and the recall intensity $\lambda = 0.02$. The expected return will be either $\mu = -0.03$ (down-market) or $\mu = 0.03$ (up-market). The initial stock price and initial entry price are taken to be $x = K = 100$. To illustrate, we focus on the put-type problem (Case (b))

and the call-type problem (Case (d)). Recall that the thresholds are given by

$$b_p = \frac{m}{m-1} aK = \left(1 + \frac{1}{m-1}\right)\left(1 - \frac{\mu}{\beta+\lambda}\right)\tilde{K}, \tag{2.11}$$

$$b_c = \frac{n}{n-1} aK = \left(1 + \frac{1}{n-1}\right)\left(1 - \frac{\mu}{\beta+\lambda}\right)\tilde{K}, \tag{2.12}$$

$$\tilde{K} = \frac{\beta}{\beta-\mu}\frac{\eta}{\rho}K = \frac{\beta-q}{\beta-\mu-\delta}K. \tag{2.13}$$

2.4.1 The impact of loan fee rate and interest rate

First, we study the impact of the loan fee rate and interest rate in the short-seller's problem. The basic role of the two parameters in the decision making by the short-seller is to determine $h_\infty(x)$, one of the two lower bounds of the value function

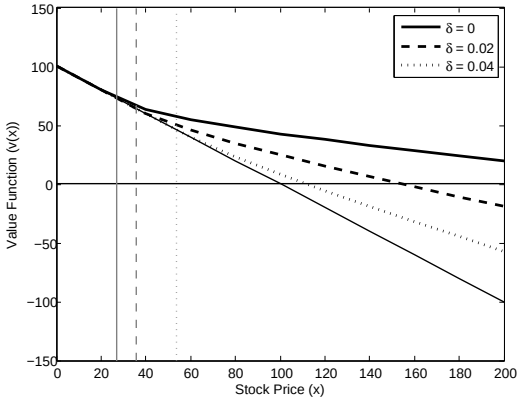
$$h_\infty(x) = \phi(x) - g(x) + K - x = \left(1 - \frac{\beta}{\beta+\lambda}\eta\right)K - \left(1 - \frac{\beta-\mu}{\beta+\lambda-\mu}\rho\right)x.$$

A different q gives a different intersection of the lower bound line via η and a different δ gives a different slope via ρ. When $q = \delta = 0$, we see $\eta = \rho = 1$, $g(x) = K - x$ and $h_\infty(x) = \phi(x)$.

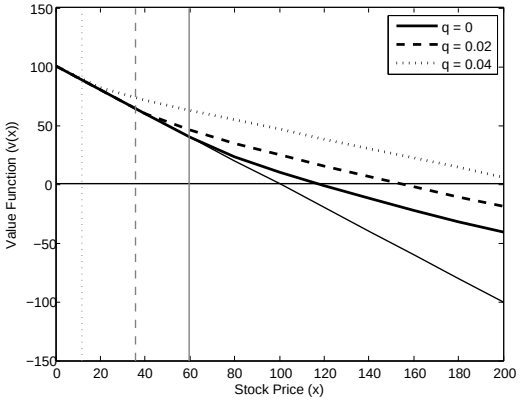
Put-type problem ($\rho > 0, \eta > 0$)

As the loan fee rate δ increases, the coefficient ρ decreases so that the slope of $h_\infty(x)$ becomes steeper while maintaining the same intersection. This generates a smaller lower bound and hence reduces the value function, as we show in Figure 2.5(a). On the other hand, as q increases and hence η decreases, the intersection of $h_\infty(x)$ increases while maintaining the same slope. Hence, the lower bound is increased and this produces a higher value function (Figure 2.5(b)). When x is close to the optimal threshold, it is less obvious by simply looking at the formula to examine the impact of the loan fee and interest income.

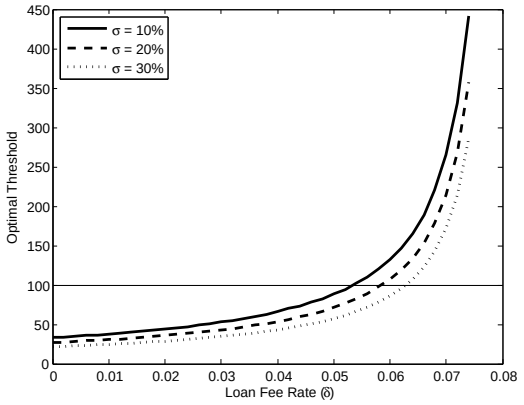
The impact of the loan fee rate and interest rate on the optimal threshold are shown in Figure 2.5(c) and 2.5(d) in a very different manner, where μ is taken as $\mu = -0.03$. The optimal threshold (2.11) is an increasing hyperbolic function of the loan fee rate δ and a decreasing linear function of the interest rate q. Figure 2.5(c) shows that the threshold is highly sensitive to the loan fee rate that is less than but close to $\beta - \mu = 0.08$. From (2.11), we see that as $\delta \uparrow \beta - \mu$ ($\rho \downarrow 0$), the threshold goes to infinity and we approach the immediate exercise solution (no entry). Figure 2.5(d) shows that the threshold depends linearly on the interest rate q. Similarly, as $q \uparrow \beta$ ($\eta \downarrow 0$), the threshold goes to zero and we recover the wait-and-see solution.



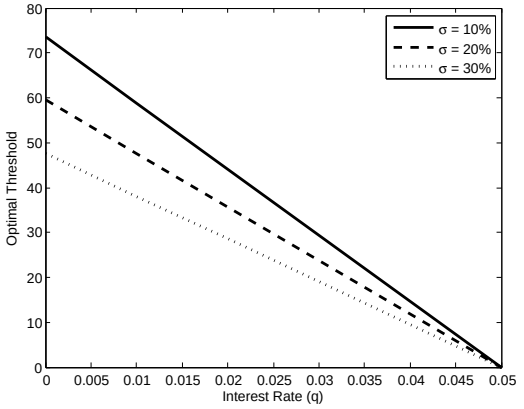
(A) Value function with loan fee rate $\delta = 0$, 0.02 and 0.04 and interest rate fixed at $q = 0.02$. The optimal thresholds are determined as 26.8, 35.7 and 53.6, respectively.



(B) Value function with interest rate $q = 0$, 0.02 and 0.04 and loan fee rate fixed at $\delta = 0.02$. The optimal thresholds are determined as 59.5, 35.7 and 11.9, respectively.



(C) Optimal threshold versus loan fee rate δ with interest rate fixed at $q = 0.02$.



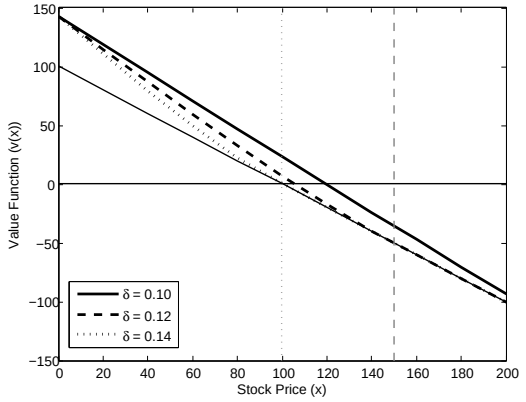
(D) Optimal threshold versus interest rate q with loan fee rate fixed at $\delta = 0.02$.

FIGURE 2.5: Value function and optimal threshold for the put-type problem.

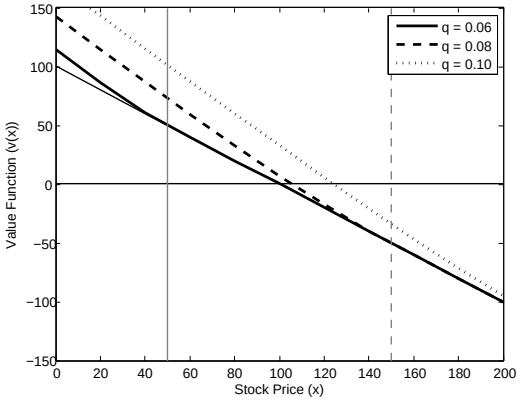
Call-type problem ($\rho < 0, \eta < 0$)

Similar analysis with varying δ and q can be applied to the call-type problem as shown in Figure 2.6(a) and 2.6(b). As the loan fee rate δ increases, the cost of holding the short position is more expensive and the investor should stop earlier, as can be seen from (2.12). Similar to the put-type problem, the optimal threshold is also highly sensitive to the loan fee (Figure 2.6(c)). From (2.12), as $\delta \downarrow \beta - \mu$ ($\rho \uparrow 0$), the threshold goes to infinity and we have the wait-and-see solution (Figure 2.6(d)).

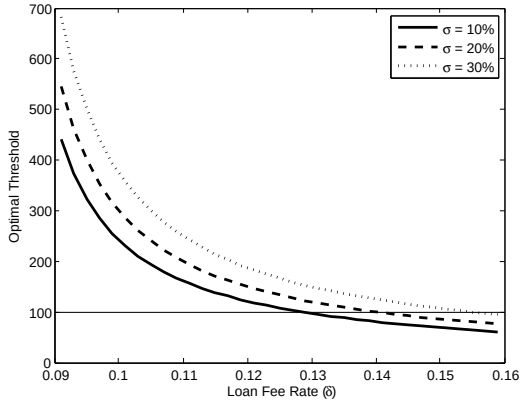
For completeness, we report in Figure 2.7 the value function in an up-market ($\mu = 0.03$) corresponding to Figures 2.5(a), 2.5(b), 2.6(a) and 2.6(b). A similar argument holds for the case of an up-market.



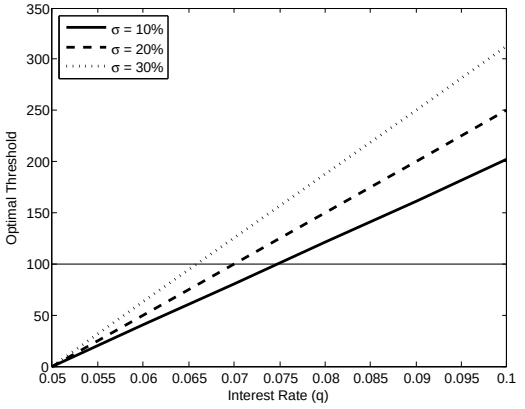
(A) Value function with loan fee rate $\delta = 0.10$, 0.12 and 0.14 and interest rate fixed at $q = 0.08$. The optimal thresholds are determined as 300, 150 and 100, respectively.



(B) Value function with interest rate $q = 0.06$, 0.08 and 0.10 and loan fee rate fixed at $\delta = 0.12$. The optimal thresholds are determined as 50, 150 and 250, respectively.



(C) Optimal threshold versus loan fee rate δ with interest rate fixed at $q = 0.08$.



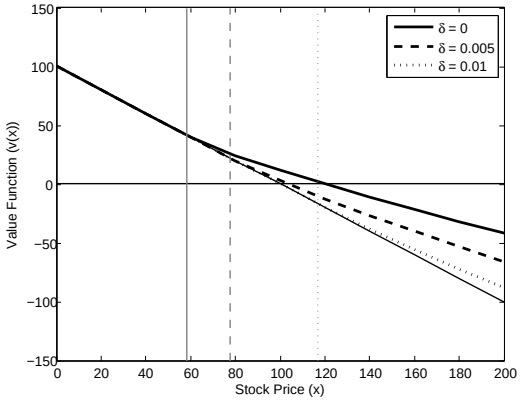
(D) Optimal threshold versus interest rate q with loan fee rate fixed at $\delta = 0.12$.

FIGURE 2.6: Value function and optimal threshold for the call-type problem.

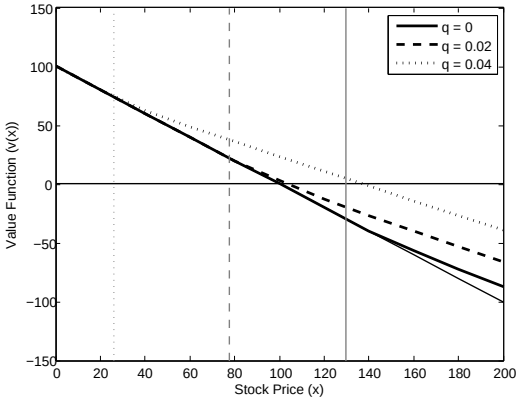
2.4.2 The impact of recall risk

The next objective is examining the impact of recall risk λ on the value function and the optimal threshold. For the time being, let us assume the simple case of $\delta = q = 0$ (equivalently, $\eta = \rho = 1$), which implies put-type problems and excludes the effects of the loan fee and interest income. We saw in the previous subsection that δ and q determine the location and the shape of the lower bound $h_\infty(x)$. Thus, from the results with $\delta = q = 0$, one can easily understand the corresponding results with nonzero δ and q by shifting vertically and rotating the line.

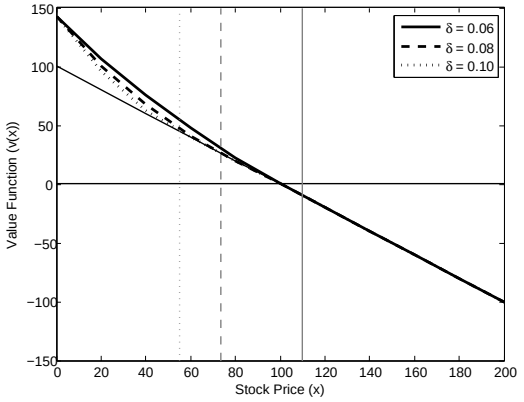
Figure 2.8(a) and 2.8(b) report the value functions $v(x)$ and the optimal thresholds b_p for different values of recall intensity λ with $\mu = -0.03$ (down-market, Figure 2.8(a)) and $\mu = 0.03$ (up-market, Figure 2.8(b)), respectively, based on Proposition 2.3 and Theorem 2.4. The corresponding optimal thresholds are calculated and marked for illustrative purposes.



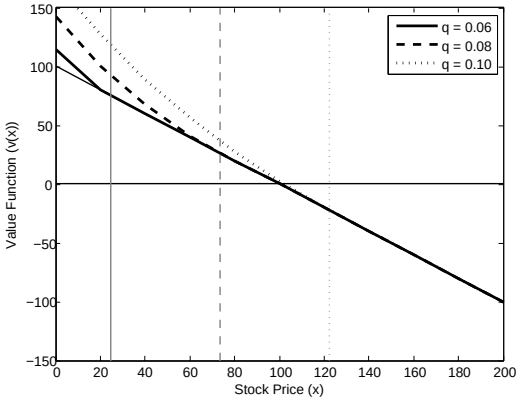
(A) Value function with loan fee rate $\delta = 0$, 0.005 and 0.01 and interest rate fixed at $q = 0.02$. The optimal thresholds are determined as 58.4, 77.9 and 116.8, respectively.



(B) Value function with interest rate $q = 0$, 0.02 and 0.04 and loan fee rate fixed at $\delta = 0.005$. The optimal thresholds are determined as 129.8, 77.9 and 26.0, respectively.



(C) Value function with loan fee rate $\delta = 0.06$, 0.08 and 0.10 and interest rate fixed at $q = 0.08$. The optimal thresholds are determined as 110, 73.4 and 55.0, respectively.



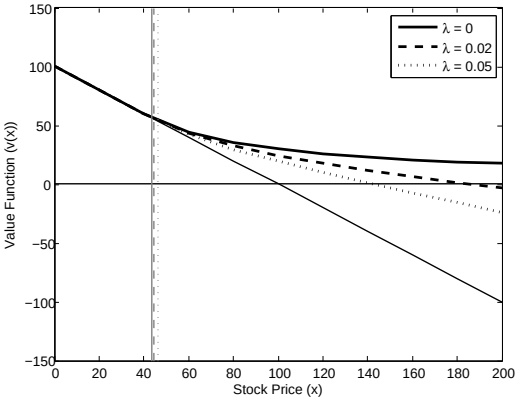
(D) Value function with interest rate $q = 0.06$, 0.08 and 0.10 and loan fee rate fixed at $\delta = 0.08$. The optimal thresholds are determined as 24.5, 73.4 and 122.3, respectively.

FIGURE 2.7: Value function in an up-market with $\mu = 0.03$.

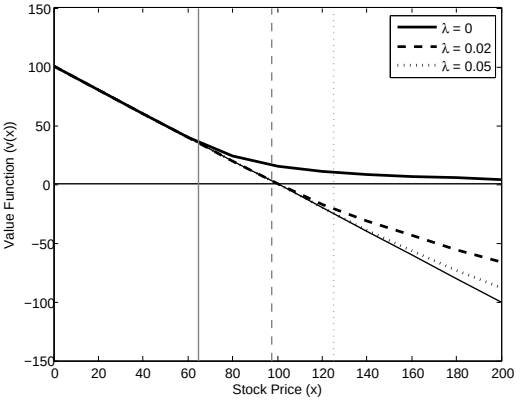
The value function $v(x)$ has two lower bounds in this case, $K - x$ and $\phi(x)$. When $\lambda = 0$, we have $\phi(x) = 0$ so that the value function is positive for all x in $\mathbb{R}_+$ (solid line).[4] As λ departs from zero, $\phi(x)$ slopes downward. Therefore, the value function can also take negative values when x is large. This is demonstrated in the cases where $\lambda = 0.02$ (dashed line) and $\lambda = 0.05$ (dotted line).

Comparing the two figures, the thresholds appear to be more sensitive to the recall risk when $\mu > 0$ (up-market) than when $\mu < 0$ (down-market). As the result the value functions do so. This phenomenon can be explained by taking a closer look at the sensitivity of b_p with respect to λ, which is positive by Proposition 2.6 (2). By

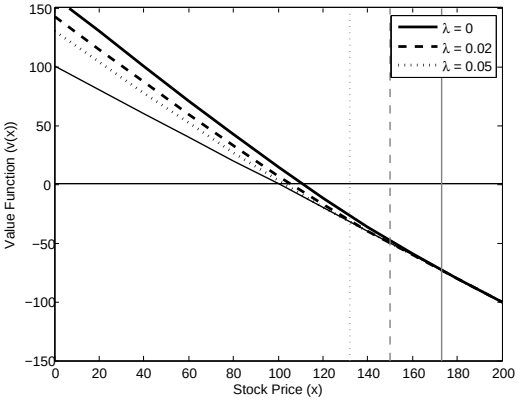
[4]This corresponds to a standard real option problem in an infinite horizon.



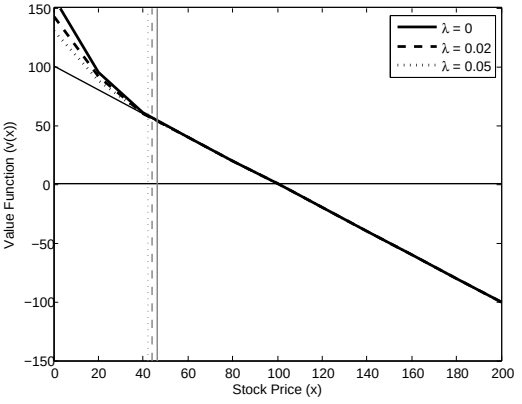
(A) Value function in a down-market ($\mu = -0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = q = 0$. The optimal thresholds are determined as 43.4, 44.6 and 46.1, respectively.



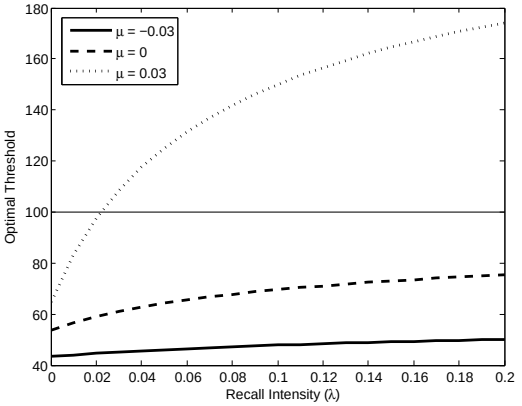
(B) Value function in an up-market ($\mu = 0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = q = 0$. The optimal thresholds are determined as 65.0, 97.3 and 125.0, respectively.



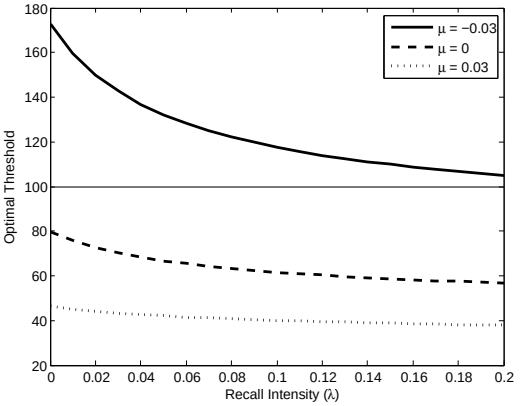
(C) Value function in a down-market ($\mu = -0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = 0.12$ and $q = 0.08$. The optimal thresholds are determined as 173.0, 150 and 132.2, respectively.



(D) Value function in an up-market ($\mu = 0.03$) with recall intensity $\lambda = 0$, 0.02 and 0.05, $\delta = 0.12$ and $q = 0.08$. The optimal thresholds are determined as 46.2, 44.0 and 42.0, respectively.



(E) Optimal threshold versus recall intensity λ with $\delta = q = 0$.



(F) Optimal threshold versus recall intensity λ with $\delta = 0.12$ and $q = 0.08$.

FIGURE 2.8: The impact of recall risk.

differentiating (2.11) with respect to λ, we see

$$\frac{\partial b_p}{\partial \lambda} = \left[\frac{\partial m}{\partial \lambda}\frac{-1}{(m-1)^2}\left(1-\frac{\mu}{\beta+\lambda}\right)+\left(1+\frac{1}{m-1}\right)\frac{\mu}{(\beta+\lambda)^2}\right)\tilde{K}.$$

As in the proof of Proposition 2.6, we have $\partial m/\partial\lambda < 0$. Hence, the first term in the brackets is always positive while the sign of the second term is the same as the sign of μ. It follows that they are cancelled out to some extent when μ is negative. This is the reason for the higher sensitivity of the thresholds for positive μ (up-market) than negative μ. Figure 2.8(e) confirms the claim: the optimal threshold increases strongly with the intensity when $\mu = 0.03$ (dotted line), while the corresponding impact is much weaker when $\mu = 0$ and $\mu = -0.03$ (dashed line and solid line, respectively). It implies that the short-seller has to be very nervous about the timing of taking capital profit (short covering) from a stock price movement with a positive trend as the recall risk increases, so that she will be satisfied with a smaller profit. On the other hand, in a down-trending market, the short-seller does not need to modify the target price so much in accordance with the magnitude of the recall risk.

As mentioned, for general δ, q (or η, ρ) leading to a put-type problem, similar analysis and interpretations hold by shifting vertically and rotating the line of the lower bound $h_\infty(x) = \phi(x) - g(x) + K - x$. From its expression, we see that the intensity λ somewhat dampens the effects of η and ρ on the lower bound $h_\infty(x)$. This is because the recall risk reduces the expected holding time of the short position and hence diminishes the role of the running cost.

For the call-type problems, these impacts are opposite to the case of the put-type problems. By differentiating (2.12) with respect to λ, we see

$$\frac{\partial b_c}{\partial \lambda} = \left[\frac{\partial n}{\partial \lambda}\frac{-1}{(n-1)^2}\left(1-\frac{\mu}{\beta+\lambda}\right)+\left(1+\frac{1}{n-1}\right)\frac{\mu}{(\beta+\lambda)^2}\right]\tilde{K},$$

which is negative by Proposition 2.6. Moreover, we have $\partial n/\partial\lambda > 0$ such that the first term in brackets is always negative while the sign of the second term is the same as μ. Hence, the sensitivity of the call-type threshold is higher for negative μ (down-market) than positive μ (up-market). Figure 2.8(f) shows that the threshold decreases notably against the intensity when $\mu = -0.03$ (solid line), while the impact is as large when $\mu = 0$ and $\mu = 0.03$ (dashed and dotted lines). As a consequence, the impact of recall risk on the value function is more significant in a down-market than in an up-market for a call-type problem (Figure 2.8(c) and 2.8(d)).

2.4.3 The impact of volatility

We then turn our attention to how the optimal threshold depends on the stock price volatility. Proposition 2.6 together with (2.11) imply that the put-type optimal threshold is decreasing with the volatility because the volatility affects b_p only through the term $m/(m-1)$ in (2.11) and $\partial m/\partial\sigma^2 > 0$. Figure 2.9(a) shows that the threshold decreases gradually with the stock price volatility under the assumption $\delta = q = 0$. The intuition is that the strategy is more opportunistic given a higher probability of the stock price declining to below the entry price. In contrast, the call-type optimal threshold is increasing with the volatility because of the term $n/(n-1)$ in (2.12) and $\partial n/\partial\sigma^2 < 0$ (Figure 2.9(c)).

Note that when $\lambda = 0$ (no recall risk) and $\delta = q = 0$, we have $b_p = (m/(m-1))K < K$ such that the investor is active regardless of the market direction. In the presence of recall risk with $\delta = q = 0$, the sign of μ matters because of

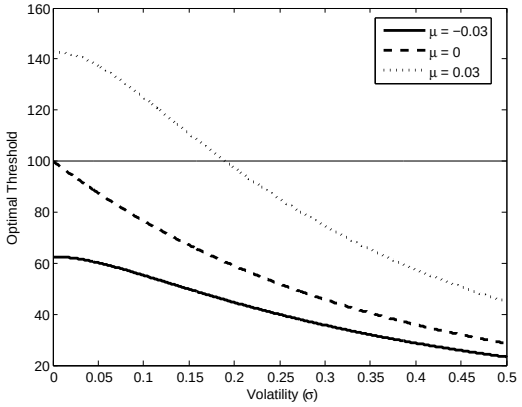
$$a = \frac{1 + \dfrac{\lambda}{\beta - \mu}}{1 + \dfrac{\lambda}{\beta}},$$

by (2.7). When $\mu \leq 0$, the investor is always active (dashed and solid lines in Figure 2.9(a)) because in this case $a \leq 1 < (m-1)/m$. In contrast, when $\mu > 0$, the investor only trades when the volatility is high enough (dotted line in Figure 2.9(a)) so that the derived value of $m < 0$ becomes large and satisfies $a < (m-1)/m$.

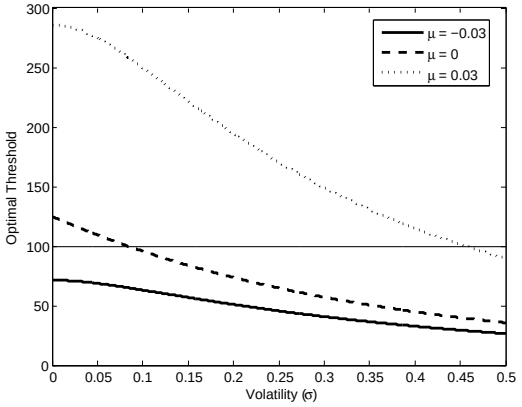
For nonzero δ, q, the condition on μ for the active region (put type or call type) can be obtained in accordance with the value of η/ρ. From (2.7), we see that the parameter a is proportional to the ratio η/ρ. For a put-type problem, keeping η/ρ and a fixed, the investor only trades when the volatility is high enough and $m < 0$ is large enough to satisfy $a < (m-1)/m$ (see Figure 2.9(a) versus 2.9(b)). Similarly, for a call-type problem, trade happens when the volatility is high enough such that $n > 1$ becomes small enough to satisfy the condition $a > (n-1)/n$ (see Figure 2.9(c) versus 2.9(d)).

2.5 Summary

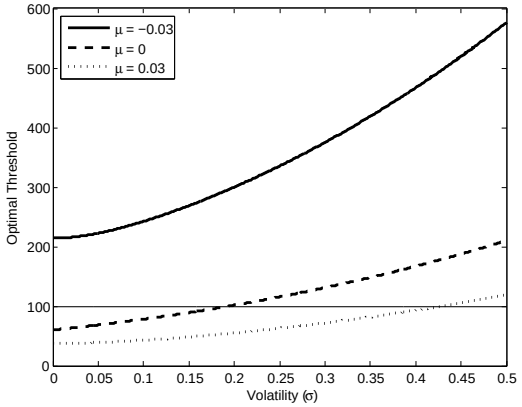
In this chapter, we studied the optimal stopping problem related to a short-selling strategy in a financial market. In a short-selling transaction, the short-seller faces the possibility of a broker recall and the short-seller might be forced to stop the strategy involuntarily and experience a loss. Our results show that, depending on the levels of the loan fee and interest rate, the optimal stopping problem is either a put-type problem



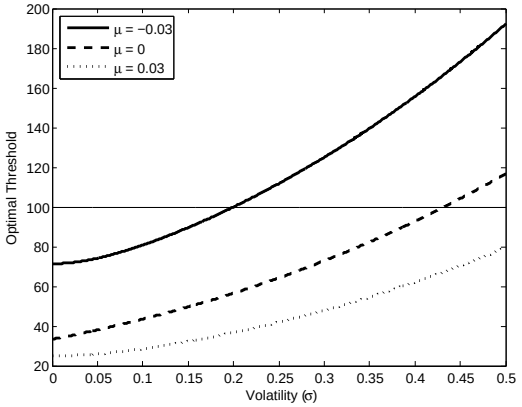
(A) Optimal threshold versus volatility σ with $\delta = q = 0$.



(B) Optimal threshold versus volatility σ with $\delta = 0.01$ and $q = 0$.



(C) Optimal threshold versus volatility σ with $\delta = 0.10$ and $q = 0.08$.



(D) Optimal threshold versus volatility σ with $\delta = 0.14$ and $q = 0.08$.

FIGURE 2.9: The impact of volatility.

with a down-and-out stopping rule or a call-type problem with an up-and-out stopping rule. The value function may become negative because of the possibility of a forced termination, and the short-seller is likely to stop earlier at the closer optimal threshold to the entry price as a result of the random recall (the put-type problem) or the relatively expensive net running cost of keeping the position (the call-type problem). The analysis in this chapter will be sufficient for investors to make a short-selling decision in a simple setting. The extension to nondiffusion-type stock price process will be discussed in the next chapter.

2.6 Appendix

2.6.1 Proof of Lemma 2.1

In this action, we bravely write the conditional expectation $\mathbb{E}[\cdot | Y_0 = x]$ as $\mathbb{E}_x[\cdot]$, which will not make confusion, in order to save the space.

(1) By the tower property and the strong Markov property, it is easy to see

$$\mathbb{E}_x\left[\int_\tau^\infty e^{-\beta s} f\left(Y_s\right) ds\right] = \mathbb{E}_x\left[e^{-\beta\tau}\mathbb{E}_{Y_\tau}\left[\int_\tau^\infty e^{-\beta(s-\tau)} f\left(Y_s\right) ds\right]\right] = \mathbb{E}_x\left[e^{-\beta\tau}\left(\mathcal{R}_\beta^Y f\right)\left(Y_\tau\right)\right].$$

This assertion is also viewed as a version of the Dynkin's formula. See p. 253 in Rogers and Williams (2000).

(2) By decomposing the event $\mathbf{1} = \mathbf{1}_{\{\tau>U\}} + \mathbf{1}_{\{\tau\le U\}}$, we see that

$$\mathbb{E}_x\left[e^{-\beta(\tau\wedge U)} f\left(Y_{\tau\wedge U}\right)\right] = \mathbb{E}_x\left[e^{-\beta U} f(X_U)\right] - \mathbb{E}_x\left[e^{-\beta U} f(X_U) 1_{\{\tau\le U\}}\right] + \mathbb{E}_x\left[e^{-\beta\tau} f(X_\tau) 1_{\{\tau\le U\}}\right].$$

The first term of the RHS is $\phi(x)$. For the second term, by the result of (1), it holds that

$$\begin{aligned}\mathbb{E}_x\left[e^{-\beta U} f(X_U) 1_{\{\tau\le U\}}\right] &= \mathbb{E}_x\left[\int_\tau^\infty e^{-\beta u} f(X_u)\lambda e^{-\lambda u} du\right] = \lambda\mathbb{E}_x\left[e^{-(\beta+\lambda)\tau}\left(R_{\beta+\lambda} f\right)\left(X_\tau\right)\right] \\ &= \mathbb{E}_x\left[e^{-(\beta+\lambda)\tau}\phi(X_\tau)\right],\end{aligned}$$

where on the last equality we used

$$\phi(x) = \mathbb{E}_x\left[e^{-\beta U} f(X_U)\right] = \mathbb{E}_x\left[\int_0^\infty e^{-\beta u} f(X_u)\lambda e^{-\lambda u} du\right] = \lambda(R_{\beta+\lambda} f)(x).$$

The third term becomes

$$\mathbb{E}_x\left[e^{-\beta\tau} f(X_\tau) 1_{\{\tau\le U\}}\right] = \mathbb{E}_x\left[e^{-\beta\tau} f(X_\tau)\int_\tau^\infty \lambda e^{-\lambda u} du\right] = \mathbb{E}_x\left[e^{-\beta\tau} f(X_\tau) e^{-\lambda\tau}\right].$$

By summing these three terms the assertion holds.

(3) The result is straight-forward for a geometric Brownian motion.

2.6.2 Short-Seller's Problem

We can simplify the short-seller's problem as follows. By Lemma 2.1 (1), we have

$$\mathbb{E}_x\left[\int_0^{\tau\wedge\tau_R} e^{-\beta s}\left(\delta X_s - qK\right) ds\right] = \left(\mathcal{R}_\beta^X f\right)(x) - \mathbb{E}_x\left[e^{-\beta(\tau\wedge\tau_R)}\left(\mathcal{R}_\beta^X f\right)\left(X_{\tau\wedge\tau_R}\right)\right].$$

Then, by Lemma 2.1 (2), the objective function of the maximization problem (2.2) is rewritten as

$$\mathbb{E}_x\left[e^{-\beta(\tau\wedge\tau_R)} g(X_\tau)\left(X_{(\tau\wedge\tau_R)}\right)\right] - \left(\mathcal{R}_\beta^X f\right)(x) = \mathbb{E}_x\left[e^{-(\beta+\lambda)\tau}\left(g-\phi\right)\left(X_\tau\right)\right] - \left(\mathcal{R}_\beta^X f\right)(x) + \phi(x).$$

The explicit expressions of g and ϕ can be obtained by Lemma 2.1 (3).

2.6.3 Proof of Proposition 2.3

Take $\alpha = \beta + \lambda$ and $h(x) = (g - \phi)(x) = kx + l$ where

$$k = -\frac{\beta - \mu}{\beta + \lambda - \mu}\rho, \quad l = \frac{\beta}{\beta + \lambda}\eta K.$$

Note that $l/k = -aK$. The stopping region can be derived as $\Gamma = F^{-1}(\hat{\Gamma})$ where

$$\hat{\Gamma} \triangleq \{y \in \mathbb{R}_+ : W(y) = H(y)\}.$$

Then,

- Case (a) When $\rho \geq 0$, $\eta \leq 0$, we have $k \leq 0$, $l \leq 0$. Hence, $\tau^* = \infty$ and $u = 0$.
- Case (b) When $\rho > 0$, $\eta > 0$, we have $k < 0$, $l > 0$. The optimal threshold is

  $$b_p = F^{-1}(y_1) = -\frac{l}{k}\frac{m}{m-1} = \frac{m}{m-1}aK.$$

  Furthermore, $H(F(b_p)) = b_p^{-m}(kb_p + l)$ and

  $$W(F(x)) = \begin{cases} x^{-m}(kx + l), & 0 \leq x \leq b_p, \\ b_p^{-m}(kb_p + l), & b_p < x. \end{cases}$$

  Hence, by Lemma 2.2, $\varphi(x)W(F(x)) = u_p(x)$ is the value function and τ^* is given by $\tau^* = \inf\{t \geq 0; X_t \leq b_p\}$.
- Case (c) When $\rho \leq 0$, $\eta \geq 0$, we have $k \geq 0$, $l \geq 0$. Hence, $\tau^* = 0$ and $u = g - \phi$. Note that $\rho = \eta = 0$ is included in Case (a) and Case (c).
- Case (d) When $\rho < 0$, $\eta < 0$, we have $k > 0$, $l < 0$. The optimal threshold is

  $$b_c = F^{-1}(y^*) = -\frac{l}{k}\frac{n}{n-1} = \frac{n}{n-1}aK,$$

  while the value function u_c can be obtained in a similar manner as in Case (b).

Chapter 3

Optimal Short-Covering with Regime Switching

3.1 Introduction

In this chapter, we extend the analysis in Chapter 2 and assume a regime-switching stock price model to capture the transition in between the bull and bear markets. This is important for the realistic modeling of a short-selling because the target price of short-covering should be sensitive to the market trend of the stock price. As a random recall with exponential distribution can be treated as a transition to an absorbing state, we are able to derive the value function and the optimal stopping rule by solving a system of differential equations along with the smooth-fit principle as in Guo and Zhang (2005).

3.2 Regime Switching Stock Price Model

3.2.1 Setup

We fix the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. The stock price is a regime-switching diffusion process $X = (X_t)_{t\geq 0}$ satisfying

$$dX_t = \mu_{J(t)} X_t dt + \sigma_{J(t)} X_t dW_t, \qquad X_0 = x > 0,$$

which is modulated by a continuous-time Markov chain $J = \{J(t)\}$ on a finite state space $E = \{1, 2\}$. Here, $\mu = \{\mu_1, \mu_2\}$ are the expected return and $\sigma = \{\sigma_1, \sigma_2\}$ with $\sigma_1 < \sigma_2$ are the discrete volatilities at each state. Furthermore, we take $\mu_1 \geq 0$, $\mu_2 < 0$

in which state 1 can be interpreted as a bull market (low volatility, upward trend) while state 2 as a bear market (high volatility, downward trend). The intensity matrix is

$$\mathbf{Q} = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix},$$

with $\lambda_i > 0$ for $i = 1, 2$. Denote

$$\mathbf{M} = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \qquad \mathbf{\Theta}(z) = \begin{pmatrix} \theta_1(z) & 0 \\ 0 & \theta_2(z) \end{pmatrix}, \qquad \mathbf{\Lambda}(z) = \begin{pmatrix} l_1(z) & \lambda_1 \\ \lambda_2 & l_2(z) \end{pmatrix},$$

where

$$\theta_i(z) = \frac{1}{2}\sigma_i^2 z^2 + \left(\mu_i - \frac{1}{2}\sigma_i^2\right) z, \qquad l_i(z) = \theta_i(z) - r - \lambda_i,$$

in which $\theta_i(z)$ is the Levy exponent corresponding to the generator

$$\mathcal{L}_i = \frac{1}{2}\sigma_i^2 x^2 \frac{d^2}{dx^2} + \mu_i x \frac{d}{dx},$$

at each state $i = 1, 2$.

The random time of the broker's recall is governed by an independent exponential random variable $\tau_R \sim Exp(\lambda_0)$ with the parameter $\lambda_0 \geq 0$. We write $\mathcal{F}_t^{W,J} = \sigma(W_s, J(s); 0 \leq s \leq t)$, $\mathbb{F}^{W,J} = \left(\mathcal{F}_t^{W,J}\right)_{t\geq 0}$ and $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, and assume that $\mathcal{F}_t^{W,J} \vee \sigma(1_{\{\tau_R > t\}}) \subset \mathcal{F}_t$. We denote the expectation

$$\mathbb{E}_{x,i}[\cdot] = \mathbb{E}[\cdot \,|\, J(0) = i, X(0) = x],$$

under $\mathbb{P}$.

In the following, we assume that the investor's own discount rate r is a constant, while the margin account interest rate and the loan fee rate are regime dependent as $\mathbf{q} = (q_1, q_2)^\top$ and $\boldsymbol{\delta} = (\delta_1, \delta_2)^\top$. The short-seller is supposed to have already undertaken the short position when the stock price was equal to K and to hold the position until she buys back at the market price either at her own discretion or following a recall by the broker. When the current state is observable as i, the short-seller's problem is to optimize the expected net profit

$$v_i(x) = \sup_{\tau_E \in \mathcal{A}} \mathbb{E}_{x,i}\left[e^{-r(\tau_E \wedge \tau_R)}\left(K - X_{\tau_E \wedge \tau_R}\right) - \int_0^{\tau_E \wedge \tau_R} e^{-rs}\left(\delta_{J(s)} X_s - q_{J(s)} K\right) ds\right], \tag{3.1}$$

for $i = 1, 2$. Here, $\mathcal{A}$ is the set of all $\mathbb{F}^{W,J}$-stopping times taking values in $[0, \infty]$. We assume that the subjective discount rate r is sufficiently high, compared with the growth rate of the stock price: $r > \max(\max(\mu_1, \mu_2), 0)$.

3.2.2 Auxiliary problem and lower bounds

Let us denote

$$\alpha_i\delta_i = \mathbf{e}_i^\top (r\mathbf{I} - \mathbf{M} - \mathbf{Q})^{-1}\boldsymbol{\delta}, \qquad \beta_i q_i = \mathbf{e}_i^\top (r\mathbf{I} - \mathbf{Q})^{-1}\mathbf{q},$$

for $i = 1, 2$. The following lemma will be useful for the calculation of the value function.

Lemma 3.1. *Suppose that* $r > \max(\max(\mu_1, \mu_2), 0)$. *Then, for any stopping time* $t > 0$ *it holds that*

$$\mathbb{E}\left[\int_t^\infty e^{-ru}\delta_{J(u)}X_u du \middle| \mathcal{F}_t\right] = e^{-rt}\alpha_{J(t)}\delta_{J(t)}X_t, \quad \mathbb{E}\left[\int_t^\infty e^{-ru}q_{J(u)}du \middle| \mathcal{F}_t\right] = e^{-rt}\beta_{J(t)}q_{J(t)},$$

and

$$\begin{aligned}
\mathbb{E}_{x,i}\left[\int_0^t e^{-ru}\delta_{J(u)}X_u du\right] &= \alpha_i\delta_i x - \mathbb{E}_{x,i}\left[e^{-rt}\alpha_{J(t)}\delta_{J(t)}X_t\right], \\
\mathbb{E}_{x,i}\left[\int_0^t e^{-ru}q_{J(u)}K du\right] &= \beta_i q_i K - \mathbb{E}_{x,i}\left[e^{-rt}\beta_{J(t)}q_{J(t)}K\right].
\end{aligned}$$

Proof. See Appendix. □

By Lemma 3.1, the short-seller's problem can be re-written as

$$v_i(x) = u_i(x) + (1 - \eta_i)K - (1 - \rho_i)x, \tag{3.2}$$

with

$$\eta_i = 1 - \beta_i q_i, \qquad \rho_i = 1 - \alpha_i\delta_i,$$

for $i = 1, 2$. The original problem is reduced to the auxiliary optimal stopping problem

$$u_i(x) = \sup_{\tau_E \in \mathcal{A}} \mathbb{E}_{x,i}\left[e^{-r(\tau_E \wedge \tau_R)}\left(\eta_{J(\tau_E\wedge\tau_R)}K - \rho_{J(\tau_E\wedge\tau_R)}X_{\tau_E\wedge\tau_R}\right)\right], \tag{3.3}$$

in which the gain function is regime dependent as

$$g_i(x) = g(x, i) = \begin{cases} \eta_1 K - \rho_1 x, & i = 1, \\ \eta_2 K - \rho_2 x, & i = 2. \end{cases} \tag{3.4}$$

The coefficients $\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$ and $\boldsymbol{\rho} = (\rho_1, \rho_2)^\top$ summarize the expected running reward and cost when the interest rate and loan fee rate are regime dependent.

The auxiliary optimal stopping problem has two lower bounds on each state i. The first one is $g_i(x)$ which corresponds to $\tau_E = 0$. The second one corresponds to $\tau_E = \infty$ as given below.

Proposition 3.2. *The lower bound of the value function $u_i(x)$ corresponding to $\tau_E = \infty$ is*

$$\psi_i(x) = a_i x + b_i, \quad i = 1, 2, \tag{3.5}$$

where

$$\mathbf{a} = \lambda_0 \left(\mathbf{M} + \mathbf{Q} - (r + \lambda_0)\mathbf{I}\right)^{-1} \boldsymbol{\rho}, \quad \mathbf{b} = \lambda_0 K \left((r + \lambda_0)\mathbf{I} - \mathbf{Q}\right)^{-1} \boldsymbol{\eta},$$

with $\mathbf{a} = (a_1, a_2)^\top$ and $\mathbf{b} = (b_1, b_2)^\top$.

Proof. See Appendix. □

The two lower bounds g_i and ψ_i determine the type of stopping rule of the optimal stopping problem at each state. When there is no regime switching, in Chapter 2 we explicitly characterize the four possible types of (i) put-type problem, (ii) call-type problem, (iii) immediate stop and (iv) wait forever, depending on the model parameters. For simplicity, we shall focus only on the put-type problem in which the stopping rule is of down-and-out type.[1] This is consistent with the general views that a short-selling strategy is to speculate a decline in the stock price. To this end, we can assume that $\mathbf{q}$ and $\boldsymbol{\delta}$ are sufficiently small as

$$\frac{q_i}{r} < 1 \quad \text{and} \quad \frac{\delta_i}{r - \mu_i} < 1 \tag{3.6}$$

in order to ensure $\eta_i > 0$ and $\rho_i > 0$ for $i = 1, 2$ (see Appendix for details). In this case, we can see that both the two lower bounds are decreasing with the stock price x, hence it is natural to conjecture a down-and-out stopping rule.

3.3 Solution

3.3.1 Value function and optimal threshold

Since $\{X_t, J_t\}$ is a joint Markov process, we can conjecture the threshold-type stopping rule for the optimal stopping problem (3.3). Following Jobert and Rogers (2006), Guo

[1] In general, there can be a switching in between the put-type and call-type problems and one needs to be more careful about the conjecture of the stopping rule.

and Zhang (2005) and Tanaka (2012), the candidate of optimal stopping time is

$$\tau_E = \min_{i=1,2} \tau_i, \qquad \tau_i = \inf\left\{t > 0; X_t \leq x_i, J\left(t\right) = i\right\}, \qquad i = 1, 2, \tag{3.7}$$

that is, the threshold is regime dependent. Let us assume the order

$$x_2 < x_1,$$

indicating that the threshold under state 2 (bear market) is lower than that under state 1 (bull market). The rationale is that the short-seller should be more aggressive and sets a lower target price for short-covering under a bear market than that of a bull market. Otherwise, we can rearrange the index and proceed similarly.

Case 1: when $x \in (0, x_2)$

The short-seller stops immediately for all regimes. We have

$$u_1(x) = \eta_1 K - \rho_1 x, \qquad u_2(x) = \eta_2 K - \rho_2 x.$$

Case 2: when $x \in [x_2, x_1)$

The short-seller stops immediately for state 1 and continues for state 2. We have

$$u_1(x) = \eta_1 K - \rho_1 x,$$

and u_2 solves the differential equation

$$\left(\mathcal{L}_2 - r\right) u_2(x) + \lambda_2 \left(u_1\left(x\right) - u_2\left(x\right)\right) + \lambda_0 \left(g_2\left(x\right) - u_2\left(x\right)\right) = 0.$$

The first term corresponds to the change in the value function u_2 without regime switching, the second term is the change when there is a regime switch from state 2 to state 1, and the third term is the change when there is a broker's random recall in state 2.

It is easy to see that the solution is

$$u_2\left(x\right) = C_1 x^{\gamma_1} + C_2 x^{\gamma_2} + \phi\left(x\right),$$

in which $\gamma_1 > 1$ and $\gamma_2 < 0$ are the roots of [2]

$$\left(r + \lambda_0 + \lambda_2\right) - \frac{1}{2}\sigma_2^2 z^2 - \left(\mu_2 - \frac{1}{2}\sigma_2^2\right) z = G_2(z) = 0,$$

[2]It can be checked that: $G_2(-\infty) < 0$, $G_2(0) > 0$, $G_2(1) > 0$ and $G_2(\infty) < 0$.

and ϕ is the special solution of the form

$$\phi(x) = \frac{\lambda_2\eta_1 + \lambda_0\eta_2}{r + \lambda_2 + \lambda_0}K - \frac{\lambda_2 p_1 + \lambda_0\rho_2}{r - \mu_2 + \lambda_2 + \lambda_0}x.$$

The coefficients C_1 and C_2 are to be determined by the smooth-fit principle.

Case 3: when $x \in [x_1, \infty)$

The short-seller continues for both the two regimes. We have to solve the system of differential equations

$$\begin{cases} (\mathcal{L}_1 - r)\, u_1(x) + \lambda_1 (u_2(x) - u_1(x)) + \lambda_0 (g_1(x) - u_1(x)) = 0 \\ (\mathcal{L}_2 - r)\, u_2(x) + \lambda_2 (u_1(x) - u_2(x)) + \lambda_0 (g_2(x) - u_2(x)) = 0 \end{cases}.$$

Define $\mathcal{A}_i = \mathcal{L}_i - (r + \lambda_0 + \lambda_i)$ and turn it into a matrix equation as

$$\begin{pmatrix} \mathcal{A}_1 & \lambda_1 \\ \lambda_2 & \mathcal{A}_2 \end{pmatrix} \begin{pmatrix} u_1(x) \\ u_2(x) \end{pmatrix} = -\lambda_0 \begin{pmatrix} \eta_1 K - \rho_1 x \\ \eta_2 K - \rho_2 x \end{pmatrix},$$

we are ready to apply the solution method in Guo and Zhang (2005). Let us focus on the homogenous equation

$$\begin{pmatrix} \mathcal{A}_1 & \lambda_1 \\ \lambda_2 & \mathcal{A}_2 \end{pmatrix} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.8}$$

We can conjecture the solution to be the linear combination of the form

$$f_1(x) = \sum_{k=1}^{4} A_k x^{\beta_k}, \qquad f_2(x) = \sum_{k=1}^{4} B_k x^{\beta_k}.$$

As shown in the Appendix, we need to solve a characteristic equation

$$G_1(\beta)\, G_2(\beta) = \lambda_1\lambda_2, \tag{3.9}$$

with

$$\begin{aligned} G_1(\beta) &= (r + \lambda_0 + \lambda_1) - \frac{1}{2}\sigma_1^2\beta^2 - \left(\mu_1 - \frac{1}{2}\sigma_1^2\right)\beta, \\ G_2(\beta) &= (r + \lambda_0 + \lambda_2) - \frac{1}{2}\sigma_2^2\beta^2 - \left(\mu_2 - \frac{1}{2}\sigma_2^2\right)\beta, \end{aligned}$$

which has the four distinct roots $\beta_4 > \beta_3 > 0 > \beta_2 > \beta_1$. Moreover, the coefficients A_k and B_k are related as

$$B_k = l_k A_k = \frac{G_1(\beta_k)}{\lambda_1} A_k = \frac{\lambda_2}{G_2(\beta_k)} A_k, \qquad k = 1, 2, 3, 4.$$

To ensure $f_1(x)$ and $f_2(x)$ are bounded as $x \to \infty$, we have to discard the positive roots β_3 and β_4. The special solution is

$$\psi_i(x) = a_i x + b_i, \qquad i = 1, 2,$$

which is just the lower bound function as obtained in Proposition 3.2. As a result, the value functions for $x \in [x_1, \infty)$ are given by

$$\begin{aligned} u_1(x) &= A_1 x^{\beta_1} + A_2 x^{\beta_2} + \psi_1(x), \\ u_2(x) &= l_1 A_1 x^{\beta_1} + l_2 A_2 x^{\beta_2} + \psi_2(x), \end{aligned}$$

in which the coefficients A_1 and A_2 are to be determined.

3.3.2 Smooth-fit

The solution for the 3 regions $(0, x_2)$, $[x_2, x_1)$ and $[x_1, \infty)$ is summarized as

$$u_1(x) = \begin{cases} \eta_1 K - \rho_1 x, & \text{if} \quad x \in (0, x_1), \\ A_1 x^{\beta_1} + A_2 x^{\beta_2} + \psi_1(x), & \text{if} \quad x \in [x_1, \infty). \end{cases} \tag{3.10}$$

$$u_2(x) = \begin{cases} \eta_2 K - \rho_2 x, & \text{if} \quad x \in (0, x_2), \\ C_1 x^{\gamma_1} + C_2 x^{\gamma_2} + \phi(x), & \text{if} \quad x \in [x_2, x_1), \\ l_1 A_1 x^{\beta_1} + l_2 A_2 x^{\beta_2} + \psi_2(x), & \text{if} \quad x \in [x_1, \infty). \end{cases} \tag{3.11}$$

Due to the exponential termination of broker's recall, we see that the special solutions appear in the value functions under all the continuation regions. The 6 unknown coefficients $A_1, A_2, C_1, C_2, x_1, x_2$ can be obtained by the smooth-fit principle as follows:

1. Matching of $u_2(x)$ at $x = x_1$:

$$\begin{aligned} l_1 A_1 x_1^{\beta_1} + l_2 A_2 x_1^{\beta_2} + \psi_2(x_1) &= C_1 x_1^{\gamma_1} + C_2 x_1^{\gamma_2} + \phi(x_1), \\ \beta_1 l_1 A_1 x_1^{\beta_1} + \beta_2 l_2 A_2 x_1^{\beta_2} + x_1 \psi_2'(x_1) &= \gamma_1 C_1 x_1^{\gamma_1} + \gamma_2 C_2 x_1^{\gamma_2} + x_1 \phi'(x_1). \end{aligned}$$

2. Matching of $u_1(x)$ at $x = x_1$:

$$\begin{aligned} A_1 x_1^{\beta_1} + A_2 x_1^{\beta_2} + \psi_1(x_1) &= \eta_1 K - \rho_1 x_1, \\ \beta_1 A_1 x_1^{\beta_1} + \beta_2 A_2 x_1^{\beta_2} + x_1 \psi_1'(x_1) &= -\rho_1 x_1. \end{aligned}$$

3. Matching of $u_2(x)$ at $x = x_2$:

$$\begin{aligned} C_1 x_2^{\gamma_1} + C_2 x_2^{\gamma_2} + \phi(x_2) &= \eta_2 K - \rho_2 x_2, \\ \gamma_1 C_1 x_2^{\gamma_1} + \gamma_2 C_2 x_2^{\gamma_2} + x_2 \phi'(x_2) &= -\rho_2 x_2. \end{aligned}$$

We can obtain a equation solving the thresholds (x_1, x_2) as

$$\begin{pmatrix} x_1^{-\gamma_1} & 0 \\ 0 & x_1^{-\gamma_2} \end{pmatrix} F_1(x_1) = \begin{pmatrix} x_2^{-\gamma_1} & 0 \\ 0 & x_2^{-\gamma_2} \end{pmatrix} F_2(x_2), \tag{3.12}$$

where $F_1(x_1)$ and $F_2(x_2)$ are given by

$$\begin{aligned} F_1(x_1) &= \begin{pmatrix} 1 & 1 \\ \gamma_1 & \gamma_2 \end{pmatrix}^{-1} \left[\begin{pmatrix} l_1 & l_2 \\ \beta_1 l_1 & \beta_2 l_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \beta_1 & \beta_2 \end{pmatrix}^{-1} \begin{pmatrix} \eta_1 K - \rho_1 x_1 - \psi_1(x_1) \\ -\rho_1 x_1 - x_1 \psi_1'(x_1) \end{pmatrix} \right. \\ &\quad \left. - \begin{pmatrix} \phi(x_1) - \psi_2(x_1) \\ x_1 \phi'(x_1) - x_1 \psi_2'(x_1) \end{pmatrix} \right], \\ F_2(x_2) &= \begin{pmatrix} 1 & 1 \\ \gamma_1 & \gamma_2 \end{pmatrix}^{-1} \begin{pmatrix} \eta_2 K - \rho_2 x_2 - \phi(x_2) \\ -\rho_2 x_2 - x_2 \phi'(x_2) \end{pmatrix}. \end{aligned}$$

Once (x_1, x_2) are obtained, we can solve (A_1, A_2) and (C_1, C_2) as

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} x_1^{\beta_1} & x_1^{\beta_2} \\ \beta_1 x_1^{\beta_1} & \beta_2 x_1^{\beta_2} \end{pmatrix}^{-1} \begin{pmatrix} \eta_1 K - \rho_1 x_1 - \psi_1(x_1) \\ -\rho_1 x_1 - x_1 \psi_1'(x_1) \end{pmatrix}, \tag{3.13}$$

$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} x_2^{\gamma_1} & x_2^{\gamma_2} \\ \gamma_1 x_2^{\gamma_1} & \gamma_2 x_2^{\gamma_2} \end{pmatrix}^{-1} \begin{pmatrix} \eta_2 K - \rho_2 x_2 - \phi(x_2) \\ -\rho_2 x_2 - x_2 \phi'(x_2) \end{pmatrix}. \tag{3.14}$$

The optimality of the value functions and thresholds can be verified by following the procedure in Guo and Zhang (2005). Lastly, it is worth to note that the derivation herein works for a two-state regime switching model only. For multiple regimes with $n > 2$, we refer the readers to Tanaka (2012) for the application of linear algebra techniques on a general regime switching model.

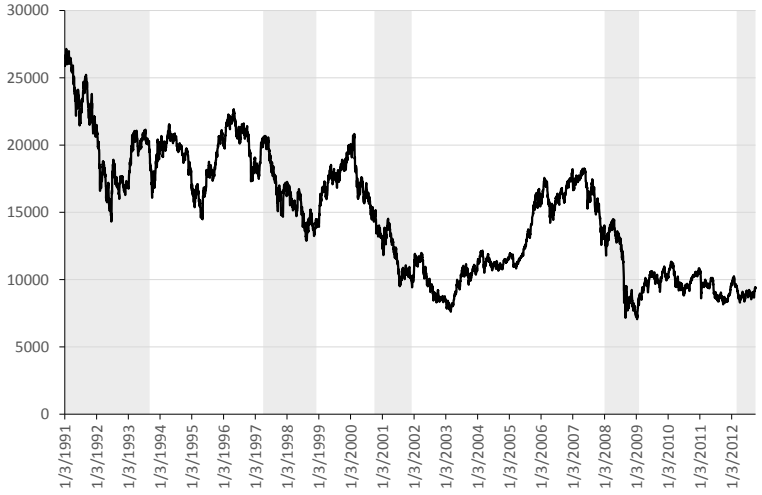


FIGURE 3.1: The Nikkei 225 index since early 1990s. The shaded area is the recession period as announced by the Economic and Social Research Institute (ESRI) in Japan.

3.4 Numerical Examples

3.4.1 Parameter calibration

We obtain the daily closing prices of the Nikkei 225 index since early 1990s and identify the bull-bear market transition based on the business cycles as announced by the Economic and Social Research Institute (ESRI) in Japan.[3] The Nikkei 225 market is an ideal venue to test our short-selling strategy because it has declined by more than 70% since the burst of the financial bubble in early 1990s. Figure 3.1 reports the Nikkei 225 index from March 1991 to November 2012 which covers several business cycles of the Japan's economy. A recession (boom) starts at the peak (trough) of a business cycle and ends at the trough (peak) as measured by the economic activity. As can be seen, the stock market declines coincide with the ESRI recession periods, suggesting that it is appropriate to determine the stock market trend based on the ESRI business cycles. To be specific, we take a boom period and a recession period to be a bull market and a bear market respectively. As such, we estimate that the average annualized log-returns during a boom period and a recession period are about 3% and -20% respectively over the sample period. Furthermore, we compute the historical volatility of Nikkei 225 using the exponentially weighted moving average (EWMA) method. As shown in Figure 3.2, the stock market volatility is usually higher during a recession period (around 30% to 40%) than that of a boom period (around 15% to 25%). This demonstrates that a regime-switching model provides a better description to the stock market dynamics than the standard Black-Scholes model with constant parameters.

[3]See http://www.esri.cao.go.jp/jp/stat/di/140530hiduke.html (in Japanese). We are grateful to the anonymous referee for the suggestion.

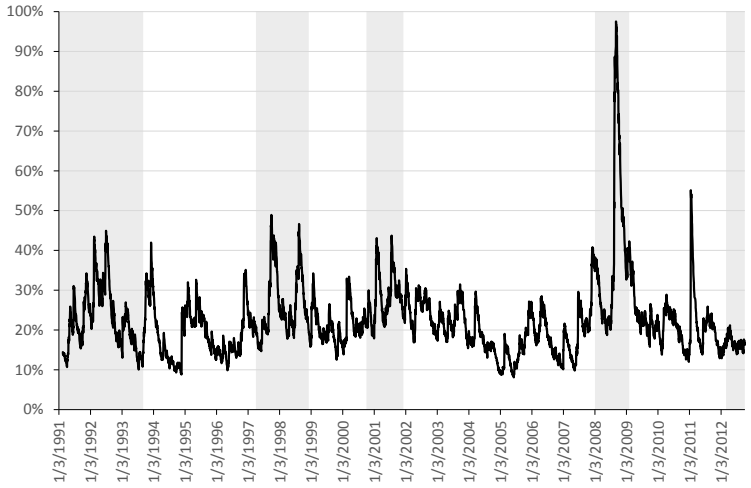


FIGURE 3.2: The historical volatility of Nikkei 225 based on the EWMA method with a decay factor of 0.06.

μ_1	σ_1	λ_1	μ_2	σ_2	λ_2
0.05	0.20	0.25	-0.10	0.40	0.75

TABLE 3.1: Model parameters for the two-state regime-switching model.

Based on the analysis, we calibrate the regime-switching model as in Table 3.1 which is explained as follows. We take state 1 to be a bull market (low volatility, upward trend) and state 2 to be a bear market (high volatility, downward trend). Conditional on the regime state i, the log-return is normally distributed as

$$\ln(X_{t+\Delta t}) - \ln(X_t) \sim N\left((\mu_i - \frac{1}{2}\sigma_i^2)\Delta t,\, \sigma_i^2 \Delta t\right),$$

where Δt is the time interval. Plugging in the values in Table 3.1, we get $\mu_1 - \sigma_1^2/2 = 0.03$ and $\mu_2 - \sigma_2^2/2 = -0.18$, which match with the empirical data. To calibrate the transition intensity, we note that on average a boom period lasts for 36 months while a recession period lasts for 16 months as indicated by the ESRI business cycles. Because the expected time to be spent on the state i is $1/\lambda_i$, the transition intensities can be chosen as

$$\lambda_1 = 0.25 \approx (1/36) \times 12, \quad \text{and} \quad \lambda_2 = 0.75 \approx (1/16) \times 12.$$

We set the recall intensity to be $\lambda_0 = 0.05$ such that a recall is a rare event as noted in D'Avolio (2002). Since the Japanese interest rates are close to zero during the sample period, we take the rebate interest rates to be $q_1 = q_2 = 0.01$ and set $\delta_1 = \delta_2 = 0.02$ for the loan fee rates. The short-seller's subjective discount rate is taken to be $r = 0.10$, while the initial stock price and entry price are normalized to be $x = K = 100$.

Figure 3.3 illustrates the value-matching and smooth-fit of the value functions v_1 and

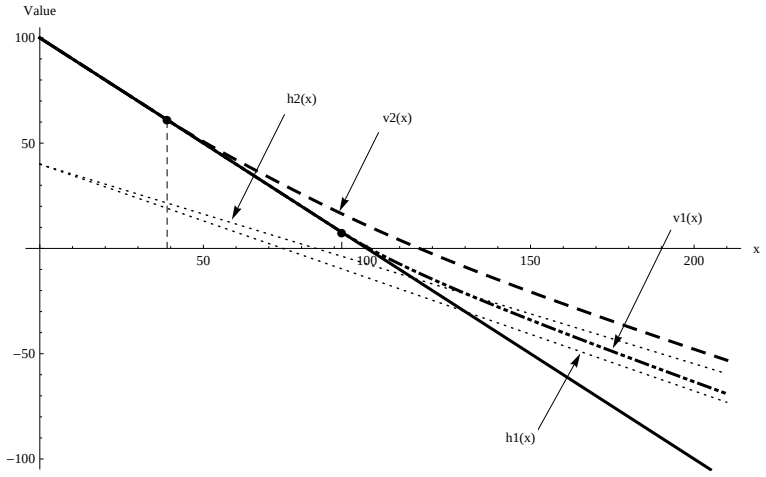


FIGURE 3.3: Value functions and optimal thresholds based on Table 3.1.

v_2 at the respective optimal thresholds $(x_1^*, x_2^*) = (92.30, 38.91)$, obtained by solving equation (3.12). The optimal threshold for short-covering in state 1 is much higher than that in state 2, suggesting that the short-seller should set a higher target price and take profit earlier under a bull market while being more aggressive in a bear market. The large difference in the optimal thresholds (x_1^*, x_2^*) highlights the importance to incorporate regime-switching into the short-seller's problem and derive the corresponding regime-dependent optimal stopping rule.

In contrast to a standard real option problem, the value functions can become negative for large x due to the broker's recall and running cost.[4] This is illustrated more clearly by plotting the lower bounds of the value functions, $h_i(x) = \psi_i(x) + (1 - \eta_i) K - (1 - \rho_i) x$ for $i = 1, 2$, which represent the expected present values of the trading strategy when the short-seller chooses to wait forever ($\tau_E = \infty$). We find that the intersection of the lower bound $h_1(x)$ and the payoff $K - x$, is quite close to x_1^* in state 1 but this is not the case in state 2. Moreover, the difference $v_2(x) - h_2(x)$ is also larger than $v_1(x) - h_1(x)$. This suggests that the optionality to stop at a finite time is more valuable in state 2 than that in state 1. Intuitively, this can be explained by the fact that a short-selling strategy is more opportunistic when the stock price tumbles in a bear market. As the value function in state 2 is quite different from that in state 1, a financial implication is that a short-seller should book the net present value of a trade (P&L) according to the current market regime.

[4]For a detailed comparative analysis for the short-seller's problem, see Chung and Tanaka (2015).

λ_0	x_1^*	x_2^*	$v_1(100)$	$v_2(100)$
0	77.68	38.04	6.25	16.08
0.02	84.37	38.51	2.74	13.12
0.05	92.30	38.91	0.59	10.87
0.1	102.25	39.30	0.00	9.16
0.2	116.06	39.77	0.00	7.71

TABLE 3.2: Impact of recall risk on the optimal thresholds and value functions.

3.4.2 The impact of recall risk

Table 3.2 reports the impact of recall risk on the optimal thresholds as the intensity λ_0 varies. It can be seen that the impact of recall risk is higher in state 1 than that in state 2. In a bull market, the stock price tends to go up and the broker's recall is more likely to lead to a trading loss. As a result, the short-seller has to be more careful about the timing of taking profit (short covering) and she needs to modify the target price accordingly as the recall risk increases. Furthermore, we note that the corresponding optimal threshold x_1^* may go over $K = 100$, indicating that the short-seller should not enter the trade in a bull market when the recall risk and stock price x are high.

3.4.3 The impact of transition intensity

We vary the transition intensities $\{\lambda_i, i = 1, 2\}$ as 0.25, 0.50, 0.75 and 1.00 to study the effect of regime-switchings to the optimal thresholds. From Table 3.3, we find that the optimal threshold under state 1 is very sensitive to the intensities λ_1 and λ_2. As the intensity λ_1 increases, there is a higher probability to switch from state 1 to state 2. This encourages the short-seller in state 1 to wait further for a transition to state 2 which is more opportunistic. When the intensity λ_2 increases, state 2 becomes relatively short-lived. The short-seller is more time constrained and has less chance to make a good profit during a bear market. Consequently, the trade in state 1 also become less opportunistic and the short-seller should cover the position earlier.

In contrast, the optimal threshold under state 2 is quite robust to the variation in transition intensities. This can be explained as follows. When the current state is 2 and $x_2^* << x_1^*$, a regime switching $(2 \to 1)$ indicates a jump from the continuation region under state 2 to the stopping region under state 1. This effectively leads to a forced termination of the strategy and such a switching is similar to a random recall. When the market is trending downward, such a forced termination is likely to lead to an early profit taking which is not too harmful to the short-seller. This is consistent with the

	λ_2			
λ_1	0.25	0.50	0.75	1.00
0.25	(77.26, 35.63)	(85.54, 37.56)	(92.30, 38.91)	(97.92, 39.91)
0.50	(63.58, 35.18)	(69.49, 37.03)	(74.73, 38.44)	(79.40, 39.52)
0.75	(57.35, 34.86)	(61.93, 36.57)	(66.13, 37.97)	(69.98, 39.09)
1.00	(53.66, 34.61)	(57.41, 36.20)	(60.90, 37.54)	(64.17, 38.67)

TABLE 3.3: Impact of transition intensity on the optimal thresholds (x_1^*, x_2^*).

analysis in Chung and Tanaka (2015) which explicitly show that the optimal threshold is less sensitive to a forced termination in a down market.

3.5 Summary

In this chapter, we study the optimal stopping problem related to a short-selling strategy in a financial market. We consider a two-state regime-switching stock price model and derive the optimal stopping rule to the short-seller's problem. When the random recall is an independent exponential variable, we are able to obtain the closed-form solution by extending the results (i.e., randomization) in Guo and Zhang (2005) and Tanaka (2012). Although we focus on the put-type problem in herein, it would be interesting to evaluate a short-selling strategy when there is a switch in between the put-type and call-type problems. Moreover, the intensity of broker's recall can also depend on the regime. These are left for future research.

3.6 Appendix

3.6.1 Proof of Lemma 3.1

Note that

$$\mathbf{\Lambda}(1) = -(r\mathbf{I} - \mathbf{M} - \mathbf{Q}), \qquad \mathbf{\Lambda}(0) = -(r\mathbf{I} - \mathbf{Q}),$$

which will be useful in the derivation. For $T \geq t$ and a real number ξ, Proposition 2.2 in Asmussen (2003) gives

$$\begin{aligned}\mathbb{E}\left[e^{-rT}\delta_{J(T)}X(T)^{\xi}\mathbf{1}_{\{J(T)=j\}}\middle| J(t)=i, X(t)\right] &= e^{-rT}X(t)^{\xi}\left\{\exp\left[\left(\mathbf{Q}+\mathbf{\Theta}(\xi)\right)(T-t)\right]\right\}_{ij} \\ &= e^{-rt}X(t)^{\xi}\left\{\exp\left[\mathbf{\Lambda}(\xi)(T-t)\right]\right\}_{ij},\end{aligned}$$

in which the second equality is due to $\mathbf{Q}+\mathbf{\Theta}(z)=\mathbf{\Lambda}(z)+r\mathbf{I}$. Applying this result, we have

$$\begin{aligned}
\mathbb{E}\left[e^{-rT}\delta_{J(T)}X(T)^{\xi}\middle|\, J(t)=i,X(t)\right] &= \sum_{j\in E}\mathbb{E}\left[e^{-rT}\delta_{J(T)}X(T)^{\xi}\mathbf{1}_{\{J(T)=j\}}\middle|\, J(t)=i,X(t)\right] \\
&= \sum_{j\in E}\mathbb{E}\left[e^{-rT}X(T)^{\xi}\mathbf{1}_{\{J(T)=j\}}\middle|\, J(t)=i,X(t)\right]\delta_j \\
&= e^{-rt}X(t)^{\xi}\sum_{j\in E}\left\{\exp\left[\mathbf{\Lambda}(\xi)(T-t)\right]\right\}_{ij}\delta_j \\
&= e^{-rt}X(t)^{\xi}\mathbf{e}_i^{\top}\exp\left[\mathbf{\Lambda}(\xi)(T-t)\right]\boldsymbol{\delta}.
\end{aligned}$$

Hence, we can compute

$$\begin{aligned}
\mathbb{E}\left[\int_t^{\infty}e^{-rt}\delta_{J(u)}X_u du\middle|\, J(t)=i,X(t)\right] &= \int_t^{\infty}\mathbb{E}\left[e^{-rt}\delta_{J(u)}X_u\middle|\, J(t)=i,X(t)\right]du \\
&= e^{-rt}X(t)\mathbf{e}_i^{\top}\int_t^{\infty}\exp\left[\mathbf{\Lambda}(1)(u-t)\right]\boldsymbol{\delta}du \\
&= e^{-rt}X(t)\mathbf{e}_i^{\top}\left\{\int_t^{\infty}\exp\left[-(r\mathbf{I}-\mathbf{M}-\mathbf{Q})(u-t)\right]du\right\}\boldsymbol{\delta} \\
&= e^{-rt}X(t)\mathbf{e}_i^{\top}(r\mathbf{I}-\mathbf{M}-\mathbf{Q})^{-1}\boldsymbol{\delta} \\
&= e^{-rt}\alpha_i\delta_i X_t,
\end{aligned}$$

in which we denote $\alpha_i\delta_i=\mathbf{e}_i^{\top}(r\mathbf{I}-\mathbf{M}-\mathbf{Q})^{-1}\boldsymbol{\delta}$. Similarly, we have

$$\begin{aligned}
\mathbb{E}_{x,i}\left[\int_t^{\infty}e^{-rt}q_{J(u)}du\middle|\, J(t)=i,X(t)\right] &= e^{-rt}\mathbf{e}_i^{\top}\int_t^{\infty}\exp\left[\mathbf{\Lambda}(0)(u-t)\right]\mathbf{q}du \\
&= e^{-rt}\mathbf{e}_i^{\top}\left\{\int_t^{\infty}\exp\left[-(r\mathbf{I}-\mathbf{Q})(u-t)\right]du\right\}\mathbf{q} \\
&= e^{-rt}\mathbf{e}_i^{\top}(r\mathbf{I}-\mathbf{Q})^{-1}\mathbf{q} \\
&= e^{-rt}\beta_i q_i,
\end{aligned}$$

in which we denote $\beta_i q_i=\mathbf{e}_i^{\top}(r\mathbf{I}-\mathbf{Q})^{-1}\mathbf{q}$.

By the strong Markov property of (X,J), we have

$$\begin{aligned}
\mathbb{E}_{x,i}\left[\int_0^{\infty}e^{-ru}\delta_{J(u)}X_u du\right] &= \alpha_i\delta_i x, \\
\mathbb{E}_{x,i}\left[\int_t^{\infty}e^{-ru}\delta_{J(u)}X_u du\right] &= \mathbb{E}_{x,i}\left[\mathbb{E}\left[\int_t^{\infty}e^{-ru}\delta_{J(u)}X_u du\middle|\,\mathcal{F}_t\right]\right]=\mathbb{E}_{x,i}\left[e^{-rt}\alpha_{J(t)}\delta_{J(t)}X_t\right],
\end{aligned}$$

such that

$$\mathbb{E}_{x,i}\left[\int_0^{t}e^{-ru}\delta_{J(u)}X_u du\right]=\alpha_i\delta_i x-\mathbb{E}_{x,i}\left[e^{-rt}\alpha_{J(t)}\delta_{J(t)}X_t\right].$$

For the second identity, we have

$$\begin{aligned}\mathbb{E}_{x,i}\left[\int_0^\infty e^{-ru}q_{J(u)}Kdu\right] &= \beta_i q_i K,\\ \mathbb{E}_{x,i}\left[\int_t^\infty e^{-ru}q_{J(u)}Kdu\right] &= \mathbb{E}_{x,i}\left[\mathbb{E}\left[\int_t^\infty e^{-ru}q_{J(u)}Kdu\middle|\mathcal{F}_t\right]\right] = \mathbb{E}_{x,i}\left[e^{-rt}\beta_{J(t)}q_{J(t)}K\right],\end{aligned}$$

such that

$$\mathbb{E}_{x,i}\left[\int_0^t e^{-ru}q_{J(u)}Kdu\right] = \beta_i q_i K - \mathbb{E}_{x,i}\left[e^{-rt}\beta_{J(t)}q_{J(t)}K\right].$$

3.6.2 Proof of Proposition 3.2

The lower bound of u_i corresponding to $\tau_E = \infty$ is

$$\begin{aligned}\psi_i(x) &= \mathbb{E}_{x,i}\left[e^{-r\tau_R}g\left(X_{\tau_R}\right)\right]\\ &= \int_0^\infty \lambda_0 e^{-(r+\lambda_0)u}\mathbb{E}_{x,i}\left[\eta_{J(u)}K\right]du - \int_0^\infty \lambda_0 e^{-(r+\lambda_0)u}\mathbb{E}_{x,i}\left[\rho_{J(u)}X_u\right]du,\end{aligned}$$

since τ_R is exponentially distributed. By Proposition 2.2 in Asmussen (2003), we have

$$\begin{aligned}\int_0^\infty \lambda_0 e^{-(r+\lambda_0)u}\mathbb{E}_{x,i}\left[\eta_{J(u)}K\right]du &= \int_0^\infty \lambda_0 K\mathbf{e}_i^\top \exp\left[-\left(\left(r+\lambda_0\right)\mathbf{I}-\mathbf{Q}\right)u\right]\boldsymbol{\eta}du\\ &= \lambda_0 K\mathbf{e}_i^\top\left(\left(r+\lambda_0\right)\mathbf{I}-\mathbf{Q}\right)^{-1}\boldsymbol{\eta},\end{aligned}$$

and

$$\begin{aligned}\int_0^\infty \lambda_0 e^{-(r+\lambda_0)u}\mathbb{E}_{x,i}\left[\rho_{J(u)}X_u\right]du &= \int_0^\infty \lambda_0 x\mathbf{e}_i^\top \exp\left[-\left(\left(r+\lambda_0\right)\mathbf{I}-\mathbf{M}-\mathbf{Q}\right)u\right]\boldsymbol{\rho}du\\ &= -\lambda_0 x\mathbf{e}_i^\top\left(\left(r+\lambda_0\right)\mathbf{I}-\mathbf{M}-\mathbf{Q}\right)^{-1}\boldsymbol{\rho}.\end{aligned}$$

Hence, we can express

$$\psi_i(x) = a_i x + b_i, \quad i = 1, 2,$$

where

$$\mathbf{a} = \lambda_0\left(\mathbf{M}+\mathbf{Q}-\left(r+\lambda_0\right)\mathbf{I}\right)^{-1}\boldsymbol{\rho}, \quad \mathbf{b} = \lambda_0 K\left(\left(r+\lambda_0\right)\mathbf{I}-\mathbf{Q}\right)^{-1}\boldsymbol{\eta},$$

with $\mathbf{a} = (a_1, a_2)^\top$ and $\mathbf{b} = (b_1, b_2)^\top$.

3.6.3 Condition (3.6)

We can compute the coefficients $\alpha_i\delta_i$ and $\beta_i q_i$ explicitly and show that the condition (3.6) implies a put-type problem with $\rho_i > 0$ and $\eta_i > 0$, for $i = 1, 2$.

1. $\alpha_i\delta_i = \mathbf{e}_i^\top (r\mathbf{I} - \mathbf{M} - \mathbf{Q})^{-1} \boldsymbol{\delta}$. Note that

$$(r\mathbf{I} - \mathbf{M} - \mathbf{Q})^{-1} = \frac{1}{\det(r\mathbf{I} - \mathbf{M} - \mathbf{Q})} \begin{pmatrix} r + \lambda_2 - \mu_2 & \lambda_1 \\ \lambda_2 & r + \lambda_1 - \mu_1 \end{pmatrix},$$

where $\det(r\mathbf{I} - \mathbf{M} - \mathbf{Q}) = (r + \lambda_1 - \mu_1)(r + \lambda_2 - \mu_2) - \lambda_1\lambda_2$. We have

$$\alpha_1\delta_1 = \frac{(r + \lambda_2 - \mu_2)\delta_1 + \lambda_1\delta_2}{(r + \lambda_1 - \mu_1)(r + \lambda_2 - \mu_2) - \lambda_1\lambda_2},$$
$$\alpha_2\delta_2 = \frac{(r + \lambda_1 - \mu_1)\delta_2 + \lambda_2\delta_1}{(r + \lambda_1 - \mu_1)(r + \lambda_2 - \mu_2) - \lambda_1\lambda_2}.$$

Under the condition $\delta_i/(r - \mu_i) < 1$ for $i = 1, 2$ from (3.6), we see that

$$\begin{aligned} 0 < (r + \lambda_2 - \mu_2)\delta_1 + \lambda_1\delta_2 &< (r + \lambda_2 - \mu_2)(r - \mu_1) + \lambda_1(r - \mu_2) \\ &= (r + \lambda_1 - \mu_1)(r + \lambda_2 - \mu_2) - \lambda_1\lambda_2, \end{aligned}$$

with $r > \max(\max(\mu_1, \mu_2), 0)$, $\lambda_i > 0$ and $\delta_i > 0$. By rearranging the terms, we can show that $\alpha_1\delta_1 < 1$ and $\rho_1 = 1 - \alpha_1\delta_1 > 0$. Similarly, we can show that $\alpha_2\delta_2 < 1$ and $\rho_2 = 1 - \alpha_2\delta_2 > 0$.

2. $\beta_i q_i = \mathbf{e}_i^\top (r\mathbf{I} - \mathbf{Q})^{-1} \mathbf{q}$. Note that

$$(r\mathbf{I} - \mathbf{Q})^{-1} = \frac{1}{\det(r\mathbf{I} - \mathbf{Q})} \begin{pmatrix} r + \lambda_2 & \lambda_1 \\ \lambda_2 & r + \lambda_1 \end{pmatrix},$$

where $\det(r\mathbf{I} - \mathbf{Q}) = (r + \lambda_1)(r + \lambda_2) - \lambda_1\lambda_2$. We have

$$\beta_1 q_1 = \frac{(r + \lambda_2)q_1 + \lambda_1 q_2}{(r + \lambda_1)(r + \lambda_2) - \lambda_1\lambda_2},$$
$$\beta_2 q_2 = \frac{(r + \lambda_1)q_2 + \lambda_2 q_1}{(r + \lambda_1)(r + \lambda_2) - \lambda_1\lambda_2}.$$

Under the condition $q_i/r < 1$ for $i = 1, 2$ from (3.6), we see that

$$\begin{aligned} 0 < (r + \lambda_2)q_1 + \lambda_1 q_2 &< (r + \lambda_2)r + \lambda_1 r \\ &= (r + \lambda_1)(r + \lambda_2) - \lambda_1\lambda_2, \end{aligned}$$

with $r > 0$, $\lambda_i > 0$ and $q_i > 0$. By rearranging the terms, we can show that $\beta_1 q_1 < 1$ and $\eta_1 = 1 - \beta_1 q_1 > 0$. Similarly, we can show that $\beta_2 q_2 < 1$ and $\eta_2 = \beta_2 q_2 > 0$.

3.6.4 Proof of (3.9)

Re-write the homogenous equation (3.8) as

$$\begin{cases} \mathcal{A}_1 f_1(x) + \lambda_1 f_2(x) = 0 \\ \mathcal{A}_2 f_2(x) + \lambda_2 f_1(x) = 0 \end{cases}.$$

We can solve $f_2(x)$ as a function of $f_1(x)$ from the first equation and substitute it into the second equation, which gives

$$\mathcal{A}_2 \mathcal{A}_1 f_1(x) = \lambda_1 \lambda_2 f_1(x).$$

Similarly, we can solve $f_1(x)$ as a function of $f_2(x)$ and obtain

$$\mathcal{A}_1 \mathcal{A}_2 f_2(x) = \lambda_1 \lambda_2 f_2(x).$$

Take the solution form

$$f_1(x) = \sum_{k=1}^{4} A_k x^{\beta_k}, \quad f_2(x) = \sum_{k=1}^{4} B_k x^{\beta_k},$$

we have

$$\begin{cases} G_1(\beta_k) G_2(\beta_k) A_k x^{\beta_k} = \lambda_1 \lambda_2 A_k x^{\beta_k} \\ G_1(\beta_k) G_2(\beta_k) B_k x^{\beta_k} = \lambda_1 \lambda_2 B_k x^{\beta_k} \end{cases},$$

which holds for $k = 1, 2, 3, 4$. Hence, we need to solve the characteristic equation

$$G_1(\beta) G_2(\beta) = \lambda_1 \lambda_2,$$

for the roots β_k with $k = 1, 2, 3, 4$. It can be shown that the four roots are distinct by checking the quartic equation

$$p(\beta) = G_1(\beta) G_2(\beta) - \lambda_1 \lambda_2,$$

which has the properties: $p(-\infty) > 0$, $p\left(\beta_1^{(1)}\right) < 0$, $p(0) > 0$, $p\left(\beta_1^{(2)}\right) < 0$, $p(\infty) > 0$, where $\beta_1^{(2)} > 1$ and $\beta_1^{(1)} < 0$ are the two distinct roots of $G_1(\beta) = 0$.

Furthermore, by a direct substitution we see that the coefficients A_k and B_k are related as

$$B_k = l_k A_k = \frac{G_1(\beta_k)}{\lambda_1} A_k = \frac{\lambda_2}{G_2(\beta_k)} A_k,$$

for $k = 1, 2, 3, 4$.

Part II

Pricing

Chapter 4

Equity-Credit Hybrid Modeling and its Application

4.1 Introduction

The affine jump-diffusion (AJD) models have been widely used in continuous time modeling of stochastic evolution of asset prices, bond yields and credit spreads. Some of the well known examples include the stochastic volatility (SV) model of Heston (1993), stochastic volatility jump-diffusion models (SVJ) of Bates (1996) and Bakshi et al. (1997), and stochastic volatility coherent jump model (SVCJ) of Duffie et al. (2000). The AJD models possess flexibility to capture the dynamics of market prices in various asset classes, while also admit nice analytical tractability. The affine term structure models, which fall into the family of AJD models, have been frequently used to study the dynamics of bond yields and credit spreads (see Duffie and Singleton, 1999).

A number of studies have addressed the importance of including jump dynamics to valuation and hedging of derivatives. In the modeling of equity derivatives, Bakshi et al. (1997) illustrate that the stochastic volatility model augmented with the jump-diffusion feature produces a parsimonious fit to stock option prices for both short-term and long-term maturities. Empirical studies reported by Bates (1996), Pan (2002) and Erakar (2004) show that the inclusion of jumps in the modeling of stock price is necessary to reconcile the time series behavior of the underlying with the cross-sectional pattern of option prices. In particular, Erakar (2004) concludes from his empirical studies that simultaneous jumps in stock price and return variance are important in catering for different volatility regimes.

While the AJD models have been successfully applied in valuation of both equity and credit derivatives, the joint modeling of equity and credit derivatives have not been fully addressed in the literature. Recently, a growing literature has highlighted such an interaction between equity risk (stock return and its variance) and credit risk (firm default risk). While the risk neutral distribution of stock return is fully conveyed by traded option prices of different strikes and maturities, the information of the arrival rate of default can be extracted from the bond yield spreads or credit default swap spreads. With the growing liquidity of the credit default swap (CDS) markets, the CDS spreads provide more reliable and updated information about the credit risk of firms. Acharya and Johnson (2007) find that the CDS market contains forward looking information on equity return, in particular during times of negative credit outlooks. For equity options, Cao et al. (2010); Cremers et al. (2008); Zhang et al. (2009) show that the out-of-the-money put options, which depict the negative tail of the underlying risk neutral distribution, are closely linked to yield spreads and CDS spreads of the reference firm.

Carr and Linetsky (2006) propose an equity-credit hybrid model in which the stock price is sent to a cemetery state upon the arrival of default of the reference company. Carr and Wu (2009) introduce another equity-credit hybrid model which incorporates jump-to-default in which the equity price drops to zero given the default arrival. Carr and Madan (2010) consider a local volatility model enhanced by jump-to-default. Mendoza-Arriaga et al. (2010) and Bayraktar and Yang (2011), respectively, propose a flexible modeling framework to unify the valuation of equity and credit derivatives using the time-changed Markov process and multiscale stochastic volatility. Cheridito and Wugalter (2012) propose a general framework under affine models with possibility of default for the simultaneous modeling of equity, government bonds, corporate bonds and derivatives.

4.2 Affine Equity-Credit Modeling

In this section, we demonstrate how to incorporate the realized variance (continuous-sampling) process in the equity-credit hybrid framework of Cheridito and Wugalter (2012). The main result is the joint moment generating function (MGF) in Proposition 4.1 which includes the realized variance as one of the Markov state variables.

4.2.1 Affine process

We fix the state space of the form $D = \mathbb{R}_+^m \times \mathbb{R}^{n-m}$ with the integers $m \geq 0$ and $n > 0$, and consider the regular affine process in Keller-Ressel et al. (2011) as:

1. $X = (X_t, \mathbb{P}_x)_{t\geq 0, x\in D}$ is a n-dimensional, time-homogenous Markov process;
2. Take the index set $\mathcal{I} = \{1, .., m\}$ and $\mathcal{J} = \{m+1, ..., n\}$ correspond to the first m-components and the last $(n-m)$-components of the Markov process X such that $X_t = \{X_{\mathcal{I},t}, X_{\mathcal{J},t}\}$.

The Markov process X is also known as the Markov state vector, and the components $\{X_{i,t}; i = 1, 2, ..n\}$ are called the Markovian factors. The first m-components of X, as denoted by $X_{\mathcal{I}}$, contains the factors that are positive. These factors are useful for the modeling of interest rate, hazard rate of default and the stochastic variance of asset price which are non-negative in nature. For example, the factors in $X_{\mathcal{I}}$ can be taken as the square-root process so as to preclude negative values (Cox et al., 1985; Heston, 1993). On the other hand, the Markovian factors in $X_{\mathcal{J}}$ allows negative values, which are useful for the modeling of the log stock price or the Guassian-type models for interest rate. These factors can be modelled by the class of Ornstein-Unlenbeck process (Vasicek, 1977).

Denote $\langle \alpha, \beta \rangle \triangleq \sum_{k=1}^{n} \alpha_k \beta_k$ as the Euclidean scalar product and $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot | x]$ as the expectation with respect to the measure $\mathbb{P}_x$. For a regular affine process, the joint moment generating function (MGF) of the Markov process X can be obtained as

$$\mathbb{E}_x \left[\exp \left(\langle \mathbf{u}, X_t \rangle \right) \right] = \exp \left(\phi(t, \mathbf{u}) + \langle \psi(t, \mathbf{u}), x \rangle \right), \tag{4.1}$$

in which the functions $\phi(t, \mathbf{u})$ and $\psi(t, \mathbf{u})$ solve a coupled system of generalized Riccati differential equation (Duffie et al., 2003).

Next, we discuss the flexibility of affine process to jointly model interest rate risk, credit risk and volatility risk in a unified manner. The key idea is that we can specify the instantaneous interest rate, hazard rate of default and stochastic variance as a linear combination of the Markovian factors.

Cumulative Interest Rate and Money Market Account

The instantaneous risk-free rate $r = (r_t)_{t\geq 0}$ is specified as

$$r_t = r_0 + \langle r_1, X_{\mathcal{I},t} \rangle, \text{ for } r_0 \in \mathbb{R}_+ \text{ and } r_1 \in \mathbb{R}_+^m, \tag{4.2}$$

which is a linear function on the positive component $X_{\mathcal{I},t}$ of the Markov process X. Furthermore, we take

$$R_t = \int_0^t r_u du, \tag{4.3}$$

as the cumulative interest rate process. The money market account $N = (N_t)_{t\geq 0}$ is given by $N_t = \exp(R_t)$, which serves as the standard numeraire for the pricing of derivatives.

It is important to note that $R = (R_t)_{t\geq 0}$ and $N = (N_t)_{t\geq 0}$ are Markov processes of finite variations.

The default process

The instantaneous hazard rate of default $\lambda = (\lambda_t)_{t\geq 0}$ is specified as

$$\lambda_t = \lambda_0 + \langle \lambda_1, X_{\mathcal{I},t} \rangle, \text{ for } \lambda_0 \in \mathbb{R}_+ \text{ and } \lambda_1 \in \mathbb{R}_+^m, \tag{4.4}$$

which is a linear function on the positive component $X_{\mathcal{I},t}$ of the Markov process X. Furthermore, we take

$$\Lambda_t = \int_0^t \lambda_u du, \tag{4.5}$$

as the cumulative hazard rate function, such that $\Lambda = (\Lambda_t)_{t\geq 0}$ is a Markov process of finite variation. With the definition of the cumulative hazard rate, we are ready to incorporate a jump-to-default feature to the affine process. The jump-to-default is generated by a doubly-stochastic process to be defined in the same state space D, with the first jump time denoted by τ as

$$\tau = \inf \{t \geq 0 : \Lambda_t \geq e\} \tag{4.6}$$

where e is a standard exponential random variable independent of X (see Lando, 1998). In the presence of default, we need to extend the state-space as $D \cup \{\Delta\}$, where Δ is a cemetery state outside of D in which the Markov process X is sent to at the time of default.

The realized variance

Lastly, the instantaneous variance process $v = (v_t)_{t\geq 0}$ is specified as

$$v_t = v_0 + \langle v_1, X_{\mathcal{I},t} \rangle, \text{ for } v_0 \in \mathbb{R}_+ \text{ and } v_1 \in \mathbb{R}_+^m, \tag{4.7}$$

which is similar to the specification of the interest rate and hazard rate processes. When there is no jump in the stock price, the continuous-monitoring realized variance is given by

$$I_t = \int_0^{t\wedge\tau} v_u du, \tag{4.8}$$

such that we evaluate the realized variance until the default time. It can be seen that $I = (I_t)_{t\geq 0}$ is a Markov process of finite variation. When the stock price has finite-activity jumps, the realized variance is modified as

$$I_t = \int_0^{t\wedge\tau} v_u du + (J_S)^2 N_{t\wedge\tau}, \tag{4.9}$$

where J_S is the jump size and N_t is an independent Poisson counter. The incorporation of jumps in the stock price and other Markov factors are possible under the general affine framework, although it might affect the analytical tractability to different extents. For exposition purpose, we shall consider the case without stock price jumps such that the realized variance is a simple integral of variance process.

The modeling of the interest rate, hazard rate and stochastic variance under the general affine framework can be summarized as

Instantaneous process	Integrated process	Examples in Modeling
$r_t = r_0 + \langle r_1, X_{\mathcal{I},t}\rangle$,	$R_t = \int_0^t r_t du$,	Money market account: $N_t = \exp(R_t)$,
$\lambda_t = \lambda_0 + \langle \lambda_1, X_{\mathcal{I},t}\rangle$,	$\Lambda_t = \int_0^t \lambda_u du$,	Default time: $\tau = \inf\{t \geq 0 : \Lambda_t \geq e\}$,
$v_t = v_0 + \langle v_1, X_{\mathcal{I},t}\rangle$,	$I_t = \int_0^t v_u du$,	Variance Swap: $\mathbb{K} = \mathbb{E}_x[I_t]$.

Hence, by choosing appropriately the loading coefficients r_1, λ_1 and v_1 one can achieve a fairly high flexibility on the joint modeling of interest rate risk, credit risk and volatility risk. Note that the incorporation of the continuous-sampling realized variance is parallel to the incorporation of stochastic interest rate and hazard rate of default.

4.2.2 Moment generating function

Cheridito and Wugalter (2012) consider an extended Markov process (X, R, Λ) that includes the cumulative interest rate $R = (R_t)_{t\geq 0}$ and cumulative hazard rate $\Lambda = (\Lambda_t)_{t\geq 0}$ as Markovian factors, and introduce the possibility of a jump-to-default in the general affine framework of Duffie et al. (2003). They derive the corresponding infinitesimal generator for the extended Markov process (which is also a regular affine process), and apply Theorem 2.7 in Duffie et al. (2003) to obtain the pre-default MGF as

$$\mathbb{E}_x\left[\exp\left(\langle \mathbf{u}, X_t\rangle + w_1 R_t + w_2 \Lambda_t\right)\mathbf{1}_{\{\tau > t\}}\right] = \exp\left(A\left(t, \mathbf{u}, w_1, w_2\right) + \langle B\left(t, \mathbf{u}, w_1, w_2\right), x\rangle\right),$$

in which $A(t, \mathbf{u}, w_1, w_2)$ and $B(t, \mathbf{u}, w_1, w_2)$ solve a system of generalized Riccati equations.

In the following, we generalize their results by including the realized variance $I = (I_t)_{t\geq 0}$ as one of the Markovian factors. We take $Y = (R, \Lambda, I) \in \mathbb{R}^3_+$ to be a Markov state vector that collects the integrated Markovian factors in (4.3), (4.5) and (4.8). It is easy to see that the extended Markov process $Z = (X, Y)$ takes value in the enlarged state space $\widetilde{D} = \mathbb{R}^m_+ \times \mathbb{R}^{n-m} \times \mathbb{R}^3_+$. Hence, Z forms an $n+3$ dimensional time-homogenous Markov process which is also a regular affine process under the definition in Keller-Ressel

et al. (2011). Following Lemma 3.1 in Cheridito and Wugalter (2012), the pre-default MGF for the extended Markov process Z is given in the next proposition.

Proposition 4.1 (Pre-default MGF)**.** *For some* $\mathbf{u} \in \mathbb{C}^n$ *and* $\mathbf{w} \in \mathbb{C}^3$*, we have*

$$\begin{aligned}\Phi_{x,t}(\mathbf{u},\mathbf{w}) &= \mathbb{E}_x\left[\exp\left(\langle \mathbf{u}, X_t\rangle + \langle \mathbf{w}, Y_t\rangle\right)\mathbf{1}_{\{\tau>t\}}\right] \\ &= \exp\left[A(t,\mathbf{u},\mathbf{w}) + \langle B(t,\mathbf{u},\mathbf{w}), x\rangle\right],\end{aligned}$$

in which $A(t,\mathbf{u},\mathbf{w})$ *and* $B(t,\mathbf{u},\mathbf{w})$ *are the solutions to the coupled system of generalized Riccati equations*

$$\begin{aligned}\partial_t A(t,\mathbf{u},\mathbf{w}) &= G_0(B(t,\mathbf{u},\mathbf{w}), w), \qquad A(0,\mathbf{u},\mathbf{w}) = 0, \\ \partial_t B_i(t,\mathbf{u},\mathbf{w}) &= G_i(B(t,\mathbf{u},\mathbf{w}), w), \qquad B_i(0,\mathbf{u},\mathbf{w}) = u_i, \\ B_j(t,\mathbf{u},\mathbf{w}) &= \exp\left(\beta_{jj}^T t\right) u_j,\end{aligned}$$

for $i \in \mathcal{I}$ *and* $j \in \mathcal{J}$*, with*

$$\begin{aligned}G_0(\xi, w) &= \langle a\xi, \xi\rangle + \langle b, \xi\rangle + r_0 w_1 + h_0(w_2 - 1) + v_0 w_3, \\ G_i(\xi, w) &= \langle \alpha_i \xi, \xi\rangle + \sum_{k=1}^{n} \langle \beta_{ki}, \xi_k\rangle + r_{1,i} w_1 + h_{1,i}(w_2 - 1) + v_{1,i} w_3,\end{aligned}$$

in which a*,* α*,* b *and* β *are the coefficients that specify the affine process and satisty the admissibility restrictions in Duffie et al. (2003).*

Proof. See Lemma 3.1 in Cheridito and Wugalter (2012). □

Note that the integrated factors $Y = (R, \Lambda, I)$ do not directly enter the MGF due to the special form of the infinitesimal generator and the fact that $Y_0 = (0, 0, 0)$. Indeed, these factors enter the system of Riccati equations through the inclusion of the constant terms $r_0 w_1 + h_0 w_2 + v_0 w_3$ for $G_0(\xi, v, w)$ and $r_{1,i} w_1 + h_{1,i} w_2 + v_{1,i} w_3$ for $G_i(\xi, v, w)$ respectively. As a result, the analytical tractability is nicely perserved with the additional the Markov factor Y.

The pre-default MGF is useful for the joint pricing of standard financial instruments such as defaultable bonds, credit default swaps and equity options. The extension herein further facilities the pricing of derivatives that are sensitive to both interest rate, credit risk and volaitlity risk. A nice example is the pricing of capped variance swap with default risk.

4.2.3 Examples

Black-Scholes model enhanced by a jump-to-default

For illustration, let us consider the Black-Scholes model enhanced by a jump-to-default

$$dS_t = S_{t_-}\left[\left(r_0 + \lambda_0\right)dt + v_0 dW_t\right],$$

where r_0 is the risk-free interest rate, λ_0 is the hazard rate of default and v_0 is the stock price volatility, which are all assumed to be constant. To put this model under the affine process framework, we take a single-factor time-homogenous Markov process $(X_t)_{t\geq 0, x\in D}$ with $D = \mathbb{R}^1$ as

$$dX_t^1 = -\frac{1}{2}v_0 dt + \sqrt{v_0}dW_t^1.$$

It is well-known that the infinitesimal generator for the Markov process X is given by $L = \frac{1}{2}\sigma^2\left[\partial_x^2 - \partial_x\right]$. Hence, we simply take $a = \frac{1}{2}\sigma^2$, $b = -\frac{1}{2}\sigma^2$ and $\alpha = \beta = 0$. The system of Riccati equation has the solution

$$A\left(t, u, \mathbf{w}\right) = \left[\frac{1}{2}u\left(u-1\right)v_0^2 + r_0 w_1 + \lambda_0\left(w_2 - 1\right) + v_0 w_3\right]t, \qquad B\left(t, u, \mathbf{w}\right) = u. \tag{4.10}$$

A three-factor hybrid model

We take a three-factor time-homogenous Markov process $X = (X_t)_{t\geq 0, x\in D}$ with the state space $D = \mathbb{R}_+^2 \times \mathbb{R}^1$ as

$$\begin{aligned} dX_t^1 &= \kappa_1\left(\theta_1 - X_t^1\right)dt + \sigma_1\sqrt{X_t^1}dW_t^1, \\ dX_t^2 &= \kappa_2\left(\theta_2 - X_t^2\right)dt + \sigma_2\sqrt{X_t^2}dW_t^2, \\ dX_t^3 &= -\frac{1}{2}X_t^1 dt + \sqrt{X_t^1}dW_t^3, \end{aligned}$$

in which $X_{\mathcal{I},t} = \left(X_t^1, X_t^2\right)$ are square-root processes and $X_{\mathcal{I},t} = \left(X_t^3\right)$. The stock price process $S = (S_t)_{t\geq 0}$ can be taken as

$$S_t = \exp\left[s_t + R_t + \Lambda_t\right]\mathbf{1}_{\{\tau > t\}},$$

in which the log stock price (s_t), stochastic volatility (ν_t) and hazard rate of default (λ_t) are to be built on the Markov process X as

$$s_t = X_t^3, \qquad \nu_t = v_0 + v_{1,1}X_t^1 + v_{1,2}X_t^2, \qquad \lambda_t = \lambda_0 + \lambda_{1,1}X_t^1 + \lambda_{1,2}X_t^2. \tag{4.11}$$

For simplicity, we assume the interest rate to be constant, i.e., $r_t = r_0$ with $r_1 = (0, 0)$. Furthermore, the correlation structure for $W = \left(W^1, W^2, W^3\right)$ is specified as

$$\begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}$$

in order to maintain analytical tractability of the affine structure.

We are ready to derive the solution for the system of generalized Riccati equations corresponding to the three-factor model. Take $a = 0$, $b = (\kappa_1\theta_1, \kappa_2\theta_2, 0)$, and

$$\alpha^1 = \frac{1}{2}\begin{pmatrix} \sigma_1^2 & 0 & \rho\sigma_1 \\ 0 & 0 & 0 \\ \rho\sigma_1 & 0 & 1 \end{pmatrix}, \quad \alpha^2 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \beta = \frac{1}{2}\begin{pmatrix} -\kappa_1 & 0 & 0 \\ 0 & -\kappa_2 & 0 \\ -1/2 & 0 & 0 \end{pmatrix}, \tag{4.12}$$

Then, we can explicitly write down the system of equations for $A(t) = A(t, \mathbf{u}, \mathbf{w})$ and $B(t) = B(t, \mathbf{u}, \mathbf{w})$ with $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ as

$$\begin{aligned}
\partial_t A(t) &= \kappa_1\theta_1 B_1(t) + \kappa_2\theta_2 B_2(t) + r_0 w_1 + \lambda_0 (w_2 - 1) + v_0 w_3, \\
\partial_t B_1(t) &= \frac{1}{2}\sigma_1^2 B_1^2(t) + (\rho\sigma_1 B_3(t) - \kappa_1) B_1(t) + \frac{1}{2} B_3(t) (B_3(t) - 1) + \lambda_{1,1} (w_2 - 1) + v_{1,1} w_3, \\
\partial_t B_2(t) &= \frac{1}{2}\sigma_2^2 B_2^2(t) - \kappa_2 B_2(t) + \lambda_{1,2} (w_2 - 1) + v_{1,2} w_3, \\
B_3(t) &= u_3,
\end{aligned}$$

for $t \geq 0$ and $A(0) = 0$ and $B(0) = (u_1, u_2, u_3)$. Note that $B_3(t) = u_3$ is a trivial solution and the two equations for $B_1(t)$ and $B_2(t)$ can be recasted as standard Riccati equations:

$$\begin{aligned}
\partial_t B_1(t) &= q_0 + q_1 B_1(t) + q_2 B_1^2(t), \\
\partial_t B_2(t) &= p_0 + p_1 B_2(t) + p_2 B_2^2(t),
\end{aligned}$$

where

$$\begin{aligned}
q_0 &= \frac{1}{2} u_3 (u_3 - 1) + \lambda_{1,1} (w_2 - 1) + v_{1,1} w_3, \qquad q_1 = \rho\sigma_1 u_3 - \kappa_1, \qquad q_2 = \frac{1}{2}\sigma_1^2, \\
p_0 &= \lambda_{1,2} (w_2 - 1) + v_{1,2} w_3, \qquad p_1 = -\kappa_2, \qquad p_2 = \frac{1}{2}\sigma_2^2,
\end{aligned}$$

with $B_1(0) = u_1$ and $B_2(0) = u_2$.

By direct integration, the system of generalized Riccati has the closed-form solution

$$\begin{aligned} A(t) &= (r_0 w_1 + \lambda_0 (w_2 - 1) + v_0 w_3)\, t \\ &\quad + \kappa_1 \theta_1 \left\{ r_- t - \frac{1}{q_2} \ln \left[\frac{1 - g \exp[-dt]}{1 - g} \right] \right\} \\ &\quad + \kappa_2 \theta_2 \left\{ s_- t - \frac{1}{p_2} \ln \left[\frac{1 - h \exp[-kt]}{1 - h} \right] \right\}, \\ B_1(t) &= \frac{r_- - r_+ g \exp[-dt]}{1 - g \exp[-dt]}, \qquad B_2(t) = \frac{s_- - s_+ h \exp[-kt]}{1 - h \exp[-kt]}, \qquad B_3(t) = u_3, \end{aligned}$$

with

$$\begin{aligned} g &= \frac{r_- - u_2}{r_+ - u_2}, \quad d = \sqrt{q_1^2 - 4 q_0 q_2}, \quad r_\pm = \frac{1}{2 q_2} [-q_1 \pm d], \\ h &= \frac{s_- - u_3}{s_+ - u_3}, \quad k = \sqrt{p_1^2 - 4 p_0 p_2}, \quad s_\pm = \frac{1}{2 p_2} [-p_1 \pm k]. \end{aligned}$$

In the case that the system of equations governing $A(t)$ and $B(t)$ has non-linear terms, we may not be able to arrive closed-form solution and one has to resort numerical method for ordinary differential equations such as the fourth-order Runga-Kutta method. Given the complicated expression of the closed-form solution, it is often useful to compute the numerical solution as a benchmark to check the implementation.

4.3 Pricing of Defaultable European Options

4.3.1 The transform analysis

Consider a defaultable European contingent claim which pays $P(X_T)$ when no default occurs before maturity and zero payoff upon default (zero recovery). Given that the payoff depends only on the terminal stock price $S_t = \exp(x_t)$, the time-t value of the contingent claim is given by

$$\begin{aligned} P(X_t, t) &= \mathbb{E}^Q \left[\exp\left(- \int_t^T r_s ds \right) P(x_T)\, \mathbf{1}_{\{\tau_d > T\}} \middle| \mathcal{F}_t \right] \\ &= \mathbf{1}_{\{\tau_d > t\}} \mathbb{E}^Q \left[\exp\left(- \int_t^T (r_s + \lambda_s) ds \right) P(x_T) \middle| \mathcal{G}_t \right], \end{aligned}$$

where $r_s + \lambda_s$ is the risk-adjusted discount rate at time s. Let $\widetilde{P}(\omega)$ denote the Fourier transform of the terminal payoff with respect to x_T, where

$$\widetilde{P}(\omega) = \int_{-\infty}^{\infty} e^{i \omega x_T} P(x_T)\, dx_T,$$

the terminal payoff can be expressed in the following representation as a generalized Fourier transform integral:

$$P\left(x_T\right)=\frac{1}{2\pi}\int_{i\varepsilon-\infty}^{i\varepsilon+\infty}e^{-i\omega x_T}\widetilde{P}\left(\omega\right)d\omega.$$

Here, the parameter $\varepsilon=\text{Im }\omega$ denotes the imaginary part of ω which falls into some regularity strip, $\varepsilon\in(a,b)$, such that the generalized Fourier transform exists Lord and Kahl (2007). By virtue of Fubini's theorem, we obtain the following integral representation of the time-t value of the contingent claim

$$P\left(X_t,t\right)=\frac{1}{2\pi}\int_{i\varepsilon-\infty}^{i\varepsilon+\infty}\Psi\left(-\omega\right)\widetilde{P}\left(\omega\right)d\omega, \tag{4.13}$$

where

$$\Psi\left(\omega\right)=\mathbf{1}_{\{\tau_d>t\}}\mathbb{E}^Q\left[\exp\left(-\int_t^T(r_s+\lambda_s)ds\right)\exp\left(i\omega x_T\right)\middle|\mathcal{G}_t\right].$$

This is just the pre-default discounted first-component marginal characteristic function. If there is a fixed recovery payment R_p to be paid on the maturity date upon earlier default, the present value of this recovery payment is given by

$$\begin{aligned}P_R\left(X_t,t\right)=R_p\Bigg\{&\mathbb{E}^Q\left[\exp\left(-\int_t^T r_s ds\right)\middle|\mathcal{G}_t\right]\\&-\mathbf{1}_{\{\tau_d>t\}}\mathbb{E}^Q\left[\exp\left(-\int_t^T(r_s+\lambda_s)ds\right)\middle|\mathcal{G}_t\right]\Bigg\}.\end{aligned} \tag{4.14}$$

All these can be readily computed using the joint MGF in Proposition 4.1.

4.3.2 European options and put-call parity

European call option

Consider a call option which pays $(S_T-K)^+$ at maturity when there is no default prior to the maturity date T and zero otherwise, so the terminal payoff function is given by

$$\left(S_T-K\right)^+\mathbf{1}_{\{\tau_d>T\}}=\left(e^{x_T}-K\right)^+\mathbf{1}_{\{\tau_d>T\}},$$

where $x_T=\ln S_T$. The Fourier transform of the above terminal payoff function is

$$\widetilde{C}\left(\omega\right)=\int_{-\infty}^{\infty}e^{i\omega x_T}\left(e^{x_T}-K\right)^+dx_T=-\frac{Ke^{i\omega\ln K}}{\omega^2-i\omega},$$

for $\varepsilon=\text{Im }\omega\in(1,\varepsilon_{\max})$. The upper bound of ε, as denoted by $\varepsilon_{\max}$, can be determined by the non-explosive moment condition $\Psi\left(-i\varepsilon\right)<\infty$ (Carr and Madan, 1999). The call

option price has the following Fourier integral representation

$$\begin{aligned} C\left(X_t,t\right) &= \frac{1}{2\pi}\int_{i\varepsilon-\infty}^{i\varepsilon+\infty}\Psi\left(-\omega\right)\left[-\frac{Ke^{i\omega\ln K}}{\omega^2-i\omega}\right]d\omega \\ &= \frac{K}{\pi}\int_0^{\infty}\mathrm{Re}\left\{\frac{e^{i(\zeta+i\varepsilon)\ln K}\Psi\left(-\left(\zeta+i\varepsilon\right)\right)}{i\left(\zeta+i\varepsilon\right)-\left(\zeta+i\varepsilon\right)^2}\right\}d\zeta,\ \omega=\zeta+i\varepsilon. \end{aligned} \tag{4.15}$$

It is worth noting that along the contour $\omega = a + ib$ for $b \in (1, \varepsilon_{\max})$, there is no singularity in the integrand and one can perform numerical integration without much difficulty.

It can be shown by replacing $k = \ln K$, $\zeta = -v$ and $\varepsilon = \alpha + 1$ that the expression in 4.15 is equivalent to the pricing formulation in Carr and Madan (1999), where

$$C\left(X_t,t\right) = \frac{e^{-\alpha k}}{\pi}\int_0^{\infty}\mathrm{Re}\left\{\frac{e^{-ivk}\Psi\left(v-i\left(\alpha+1\right)\right)}{-\left(v-i\alpha\right)\left[v-i\left(\alpha+1\right)\right]}\right\}dv.$$

In other words, one would obtain the same analytic expression of the Fourier integral representation no matter whether one considers the transform with respect to the log-stock price or log-strike price.

European put option

Suppose a put option pays at maturity the following payoff: $\left(K-S_T\right)^+\mathbf{1}_{\{\tau_d>T\}} = \left(K-e^{x_T}\right)^+\mathbf{1}_{\{\tau_d>T\}}$ when there is no default prior to maturity, and a recovery payment $R_P\mathbf{1}_{\{\tau_d\le T\}}$ to be paid at maturity when default occurs during the contractual period. The Fourier transform of the payoff of the non-default component of the put option is given by

$$\widetilde{P}_0\left(\omega\right) = \int_{-\infty}^{\infty}e^{i\omega x_T}\left(K-e^{x_T}\right)^+dx_T = -\frac{Ke^{i\omega\ln K}}{\omega^2-i\omega},$$

for $\varepsilon = \mathrm{Im}\ \omega \in (-\varepsilon_{\max}, 0)$. Inside the regularity strip where the above Fourier transform is well defined, the non-default component has the following integral representation

$$\begin{aligned} P_0\left(X_t,t\right) &= \frac{1}{2\pi}\int_{i\varepsilon-\infty}^{i\varepsilon+\infty}\Psi\left(-\omega\right)\left(-\frac{Ke^{i\omega\ln K}}{\omega^2-i\omega}\right)d\omega \\ &= \frac{K}{\pi}\int_0^{\infty}\mathrm{Re}\left\{\frac{e^{i(\zeta+i\varepsilon)\ln K}\Psi\left(-\left(\zeta+i\varepsilon\right)\right)}{i\left(\zeta+i\varepsilon\right)-\left(\zeta+i\varepsilon\right)^2}\right\}d\zeta,\ \omega=\zeta+i\varepsilon. \end{aligned}$$

It is interesting to find that the Fourier transform of the terminal payoff function for the put option and the call option counterpart both have the same integral representation, though subject to different constraints on the regularity strip. Note that $\mathrm{Im}\ \omega \in (1, \varepsilon_{\max})$ for the call option and $\mathrm{Im}\ \omega \in (-\varepsilon_{\max}, 0)$ for the put option.

The recovery payment can be obtained similar to $P_R(X_t, t)$ in the evaluation of a defaultable bond. The defaultable European put option price is then given by

$$P(X_t, t) = P_0(X_t, t) + P_R(X_t, t).$$

Put-call parity relation under jump-to-default

Take the recovery payment to be $R_p = K$. In the presence of jump-to-default, a portfolio of a long call and a short put has the terminal payoff

$$\begin{aligned} & (S_T - K)^+ \mathbf{1}_{\{\tau_d > T\}} - \left[(K - S_T)^+ \mathbf{1}_{\{\tau_d > T\}} + K\mathbf{1}_{\{\tau_d \leq T\}}\right] \\ = \; & (S_T - K)\mathbf{1}_{\{\tau_d > T\}} - K\mathbf{1}_{\{\tau_d \leq T\}}. \end{aligned}$$

Hence, the difference of defaultable European call and put prices is given by

$$C(X_t, t) - P(X_t, t) = \mathbf{1}_{\{\tau_d > t\}} S_t - KB_f(t, T),$$

where $B_f(t, T)$ is defined before. The put-call parity relation in the presence of jump-to-default is seen to be the same as the standard relation. This is consistent with the model-free property of the put-call parity relation.

4.4 Pricing of Capped Variance Swaps

Denote the contractual period of the variance swap to be $[0, T]$ with the monitoring dates $0 < t_1 < ... < t_i ... < t_N = T$, where N is the total number of monitoring dates. Denote K to be fair strike of the variance swap contract that equates its initial entry cost to be zero. It is well-known that the pricing of variance swap is related to the realized variance of the stock price over the contractual period as

$$I(0, T; N) = \frac{A}{N} \sum_{i=1}^{N} \left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2, \tag{4.16}$$

where S_t is the stock price and $A = N/T$ is the annualized factor. When there is no default risk, the payoff of a variance swap at maturity is given by $I(0, T; N) - K$ such that the fair strike is $K = \mathbb{E}_x[I(0, T; N)]$.

Denote τ to be the default time, $\widetilde{S} = \left(\widetilde{S}_t\right)_{0\leq t\leq \tau}$ to be the pre-default stock price process and the cemetery state asΔ. The defaultable stock price process can be defined as

$$S_t = \begin{cases} \widetilde{S}_t, & \tau > t, \\ \Delta, & \tau \leq t. \end{cases}$$

Suppose that the default time occurs in between two monitoring dates such that $\tau \in (t_{n-1}, t_n]$ for $n = 1, 2, ...N$, we can decompose the realized variance into three terms as

$$\begin{aligned} \sum_{i=1}^{N} \left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 &= \sum_{i=1}^{n-1} \left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 + \left[\ln\left(S_{t_n}/S_{t_{n-1}}\right)\right]^2 + \sum_{i=n+1}^{N} \left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 \\ &= \sum_{i=1}^{n-1} \left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 + \left[\ln\left(\Delta/S_{t_{n-1}}\right)\right]^2 \end{aligned}$$

in which we set the last term to zero because $S_{t_{i-1}} = \Delta$ for $i = n+1, ..., N$. Furthermore, let us send Δ arbitrarily close to zero, the log-return $\ln\left(\Delta/S_{t_{n-1}}\right)$ and corresponding squared-return become infinite due to the default time τ. In view of this, we have to formulate

$$I(0, T; N) = \begin{cases} \frac{A}{N} \sum_{i=1}^{N} \left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2, & \tau > T, \\ \infty, & \tau \leq T, \end{cases} \tag{4.17}$$

as the defaultable realized variance. Next, we discuss the pricing of capped variance swap based on the defaultable realized variance.

4.4.1 The cap feature

In the aftermath of the Lehman financial crisis, it has become standard to set a cap to single-name variance swap to limit the risk exposure of the underwriter (Alexander and Leontsinis, 2012). For a capped variance swap, the payoff $p_T(K)$ at the maturity is given by

$$p_T(K) = \min\left[I(0, T; N), c\right] - K \tag{4.18}$$

in which $I(0, T; N)$ is the defaultable realized variance, and the cap is specified as

$$c = m_{cap} K,$$

in which m_{cap} is a pre-determined multiplier that is usually set to be 2.5^2 times of the fair strike. Decompose the default event $\mathbf{1} = \mathbf{1}_{\{\tau\leq T\}} + \mathbf{1}_{\{\tau>T\}}$, we see that

$$\begin{aligned} \min\left[I(0, T; N), c\right] &= \min\left[I(0, T; N), c\right] \mathbf{1}_{\{\tau>T\}} + \min\left[I(0, T; N), c\right] \mathbf{1}_{\{\tau\leq T\}} \\ &= \min\left[I(0, T; N), c\right] \mathbf{1}_{\{\tau>T\}} + c\mathbf{1}_{\{\tau\leq T\}}, \end{aligned}$$

in which the second identity follows from the fact that the cap is hit immediately upon the default event due to (4.17). Furthermore, apply the identity $\min(a,b) = \min(a-b,0)+b$, we can express the first term as

$$\min\left[I\left(0,T;N\right),c\right]\mathbf{1}_{\{\tau>T\}} = \min\left[I\left(0,T;N\right)-c,0\right]\mathbf{1}_{\{\tau>T\}} + c\mathbf{1}_{\{\tau>T\}},$$

such that

$$p_T\left(K\right) = \min\left[I\left(0,T;N\right)-c,0\right]\mathbf{1}_{\{\tau>T\}} + \left(c-K\right).$$

The payoff of a capped variance swap can be splitted into the following 3 cases:

1. When there is no default (i.e. $\mathbf{1}_{\{\tau>T\}}=1$) and $I\left(0,T;N\right)\leq c$, we have

   $$p_T\left(K\right) = I\left(0,T;N\right)-K,$$

   which is the payoff of a standard variance swap.

2. When there is no default (i.e. $\mathbf{1}_{\{\tau>T\}}=1$) and $I\left(0,T;N\right)>c$, we have

   $$p_T\left(K\right) = \left(c-K\right),$$

   that is, the capped payoff is attained due to the accumulation of realized variance.

3. When there is default (i.e. $\mathbf{1}_{\{\tau>T\}}=0$), we have $I\left(0,T;N\right)=\infty$ and hence

   $$p_T\left(K\right) = \left(c-K\right),$$

   that is, the capped payoff is attained due to explosion in the realized variance at a default.

Alternatively, we can rewrite

$$p_T\left(K\right) = \left(c-K\right) - \max\left[c-I\left(0,T;N\right),0\right]\mathbf{1}_{\{\tau>T\}},$$

by $\min\left(-a,0\right) = -\max\left(a,0\right)$. This shows that the pricing of a capped variance swap can be decomposed into a constant payoff $c-K$ plus a correction term which is related to a defaultable variance put option with strike c. Buehler (2010) and Cheridito and Wugalter (2012) assume that the cap is hit only at the time of default. Chen (2012) discusses the pricing of capped volatility swap under Heston model and derive a closed-form formula using Laplace transform of the continuous-monitoring realized variance.

4.4.2 Continuous-monitoring

Take $N \to \infty$, on the event $\{\tau > T\}$ the continuous-monitoring limit of realized variance is given by

$$\lim_{N\to\infty} I(0,T;N) = I_T = \int_0^T v_u du,$$

such that

$$p_T(K) = (c-K) - \max[c - I_T, 0]\,\mathbf{1}_{\{\tau>T\}}, \tag{4.19}$$

with I_t is the continuous-monitoring realized variance as defined in (4.8). Thanks to the analytical tractability for I_t as one of the Markovian factor in the affine process, it is possible to derive the formula for the capped variance swap using the marginal MGF of I_t.

Marginal MGF and variance put

From the joint MGF $\Phi_{x,t}(u,w)$ for the extended Markov process $Z = (X,Y)$, the marginal MGF $\Phi_t^I(\cdot)$ for the realized variance I can be obtained by taking $u = \mathbf{0}_n = (0,,...,0)$ and $w = (0,0,\zeta)$ as

$$\mathbb{E}_x\left[\exp(\zeta I_t)\,\mathbf{1}_{\{\tau>t\}}\right] = \Phi_{x,t}(\mathbf{0}_n, (0,0,\zeta)). \tag{4.20}$$

Make use of the integral representations for $(c-x)^+ = \max[c-x,0]$ as

$$(c-x)^+ = \frac{1}{2\pi i}\int_0^\infty \frac{e^{-cz}}{z^2}e^{zx}dz = \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left\{\frac{e^{-c(\varepsilon+iy)}}{(\varepsilon+iy)^2}e^{(\varepsilon+iy)x}\right\}dy,$$

for $x \geq 0$ and $\varepsilon < 0$. The defaultable variance put can be obtained as (Kallsen et al., 2010)

$$\mathbb{E}_x\left[(c-I_T)^+\,\mathbf{1}_{\{\tau>T\}}\right] = \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left\{\frac{e^{-c(\varepsilon+iy)}}{(\varepsilon+iy)^2}\Phi_{x,t}(\mathbf{0}_n,(0,0,\varepsilon+iy))\right\}dy, \tag{4.21}$$

in which T is the maturity of the option.

Probability of hitting the cap

Buehler (2010) and Cheridito and Wugalter (2012) assume that the cap is attained only at the time of jump-to-default. This gives the approximation

$$p_T(K) = [I(0,T;N) - K]\,\mathbf{1}_{\{\tau>T\}} + c\mathbf{1}_{\{\tau\leq T\}}, \tag{4.22}$$

which ignores the probability of hitting the cap before the occurence of a jump-to-default. The validity of the approximation is equivalent to assuming that the following excess probability is close to zero:

$$\Pr\left\{I\left(0,T;N\right)\geq c\right\}\approx 0.$$

In the continuous-monitoring limit, it is possible to compute the excess probability explicitly using the transform approach. From the marginal MGF for I_t

$$\mathbb{E}_x\left[\exp\left(\zeta I_t\right)\mathbf{1}_{\{\tau>t\}}\right]=\Phi_{t,x}\left(\mathbf{0}_n,(0,0,\zeta)\right), \tag{4.23}$$

we can compute the excess probability as

$$\Pr\left\{I_T\geq c\right\}=\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-iyc}\Phi_{x,t}\left(\mathbf{0}_n,(0,0,iy)\right)}{iy}\right]dy,$$

based on the well-known Levy inversion formula (Gil-Pelaez, 1951).

4.4.3 Discrete-monitoring

We would like to explore the analytical tractability in the case of discrete-monitoring. It is known that the pricing of variance option with discrete-monitoring requires one to resort to approximations or numerical methods. To this end, let us assume that the excess probability of the realized variance over the cap $c=m_{cap}K$ to be negligible as

$$\Pr\left\{I\left(0,T;N\right)\geq c\right\}\approx 0,$$

such that we can approximate the payoff of a capped variance swap using

$$p_T\left(K\right)=\left[I\left(0,T;N\right)-K\right]\mathbf{1}_{\{\tau>T\}}+c\mathbf{1}_{\{\tau\leq T\}}.$$

In that case, we have

$$\mathbb{E}_x\left[I\left(0,T;N\right)\mathbf{1}_{\{\tau>T\}}\right]=K\left(\mathbb{E}_x\left[\mathbf{1}_{\{\tau>T\}}\right]-m_{cap}\mathbb{E}_x\left[\mathbf{1}_{\{\tau\leq T\}}\right]\right) \tag{4.24}$$

such that the fair strike K of discrete-monitoring capped variance swap is obtained as

$$K=\frac{\mathbb{E}_x\left[I\left(0,T;N\right)\mathbf{1}_{\{\tau>T\}}\right]}{\mathbb{E}_x\left[\mathbf{1}_{\{\tau>T\}}\right]-m_{cap}\mathbb{E}_x\left[\mathbf{1}_{\{\tau\leq T\}}\right]}. \tag{4.25}$$

From the expression, it is suffice to evaluate of the expectation with respect to the pre-default realized variance as $\mathbb{E}_x\left[I\left(0,T;N\right)\mathbf{1}_{\{\tau>T\}}\right]$. To this end, we see that

$$\mathbb{E}_x\left[I\left(0,T;N\right)\mathbf{1}_{\{\tau>T\}}\right]=\frac{A}{N}\sum_{i=1}^{N}\mathbb{E}_x\left[\left(\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right)^2\mathbf{1}_{\{\tau>T\}}\right]=\frac{A}{N}\sum_{i=1}^{N}\rho\left(t_{i-1},t_i;T\right),$$

in which we take

$$\rho\left(t_{i-1},t_i;T\right)=\mathbb{E}_x\left[\left(\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right)^2\mathbf{1}_{\{\tau>T\}}\right] \tag{4.26}$$

as the pre-default squared-return.

Let us further specify the stock price process using the Markov process X. We construct the risk-neutral dynamics of the defaultable stock price process $S=(S_t)_{t\geq 0}$ as

$$\begin{aligned} S_t &= \exp\left[\xi_t+R_t+\Lambda_t\right]\mathbf{1}_{\{\tau>t\}}, \\ \xi_t &= \left\langle\delta_1,X_t\right\rangle, \quad \text{for} \quad \delta_1\in\mathbb{R}^n, \end{aligned}$$

in which $\xi=(\xi_t)_{t\geq 0}$ is the factor driving the stock price process. We further define the log stock price process $s=(s_t)_{t\geq 0}$

$$s_t=\xi_t+R_t+\Lambda_t,$$

which will be useful for the derivation.

Now, we are ready to the pricing of capped variance swap with discrete-monitoring. In the following, we take $\mathbf{0}_n$ and $\mathbf{0}_3$ to denote a zero vector of the dimension n and 3 respectively. The functions $A\left(t,u,w\right)$ and $B\left(t,u,w\right)$ should be known as the solution to the system of generalized Riccati equations in (4.10).

Lemma 4.2. *For any $t\in[0,T]$ and any measurable function π_t with respect to $\sigma\left(X_u|\,u\leq t\right)$, then*

$$\mathbb{E}_x\left[\pi_t\mathbf{1}_{\{\tau>T\}}\right]=\mathbb{E}_x\left[d_{t,T}\pi_t\mathbf{1}_{\{\tau>t\}}\right],$$

where $d_{t,T}$ is given by

$$d_{t,T}=\mathbb{E}\left[\mathbf{1}_{\{\tau>T\}}\middle|\,X_t,\mathbf{1}_{\{\tau>t\}}\right]=\exp\left[A\left(T-t,\mathbf{0}_n,\mathbf{0}_3\right)+\left\langle B\left(T-t,\mathbf{0}_n,\mathbf{0}_3\right),X_t\right\rangle\right],$$

with $A\left(t,u,w\right)$ and $B\left(t,u,w\right)$ are the solution to the system of generalized Riccati equations in (4.10).

Proof. See Appendix. □

To proceed, we resort to the identity from Zheng and Kwok (2014) as

$$\left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 = \frac{\partial^2}{\partial\phi^2}\exp\left[\phi\left(s_{t_i}-s_{t_{i-1}}\right)\right]\Big|_{\phi=0}.$$

By Lemma 4.2 and the Fubini's theorem, we can express the pre-default squared-return as

$$\rho\left(t_{i-1},t_i;T\right) = \mathbb{E}_x\left[\left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2\mathbf{1}_{\{\tau>T\}}\right] = \frac{\partial^2}{\partial\phi^2}\Psi_x\left(\phi;t_{i-1},t_i,T\right)\Big|_{\phi=0}$$

where

$$\Psi_x\left(\phi;t_{i-1},t_i,T\right) = \mathbb{E}_x\left[\exp\left(A_1+\langle B_1,X_{t_i}\rangle\right)\exp\left(\phi\left(s_{t_i}-s_{t_{i-1}}\right)\right)\mathbf{1}_{\{\tau>t_i\}}\right]$$

with A_1 and B_1 given in the next Theorem.

Theorem 4.3 (Pre-default Squared-Return)**.** *For $0 \le t_{i-1} < t_i \le T$, the pre-default squared-return in (4.26) can be evaluated as*

$$\rho\left(t_{i-1},t_i;T\right) = \frac{\partial^2}{\partial\phi^2}\Psi_x\left(\phi;t_{i-1},t_i,T\right)\Big|_{\phi=0}, \tag{4.27}$$

where

$$\Psi_x\left(\phi;t_{i-1},t_i,T\right) = \exp\left(\sum_{i=1}^{3}A_i+\langle B_3,x\rangle\right),$$

with

$$\begin{aligned} A_1 &= A\left(T-t_i,\mathbf{0}_n,\mathbf{0}_3\right), \quad A_2 = A\left(t_i-t_{i-1},B_1+\phi\delta_1,\mathbf{w}_\phi\right), \quad A_3 = A\left(t_{i-1},B_2-\phi\delta_1,\mathbf{0}_3\right) \\ B_3 &= B\left(t_{i-1},B_2-\phi\delta_1,\mathbf{0}_3\right), \quad B_2 = B\left(t_i-t_{i-1},B_1+\phi\delta_1,\mathbf{w}_\phi\right), \quad B_1 = B\left(T-t,\mathbf{0}_n,\mathbf{0}_3\right), \end{aligned}$$

in which $A\left(t,\mathbf{u},\mathbf{w}\right)$ and $B\left(t,\mathbf{u},\mathbf{w}\right)$ are solution to the Riccati equations in Proposition 4.1.

Proof. See Appendix. □

We note that the computation of the pre-default squared-return $\rho\left(t_{i-1},t_i;T\right)$ is related to three random variables: $S\left(t_i\right), S\left(t_{i-1}\right)$ and $\mathbf{1}_{\{\tau>T\}}$. The calculation of the successive iterated expectations is made possible given the exponential form of the MGF under affine models.

Denote the default probability as $p = \mathbb{E}_x\left[\mathbf{1}_{\{\tau\le T\}}\right]$. When the condition $m_{cap} < (1/p-1)$ holds, the fair strike of discrete-monitoring capped variance swap can be approximated

as

$$K = \left[\frac{A}{N}\sum_{i=1}^{N}\rho\left(t_{i-1}, t_i; T\right)\right]\frac{1}{\gamma\left(x, T\right)} \tag{4.28}$$

where $\gamma = 1 - (1 + m_{cap})p$. It is interesting to note the properties of the formula:

1. Denote p as the default probability. Evidently, we need to impose the restriction $m_{cap} < (1/p - 1)$, for $0 \leq p < 1$, in order to keep the fair strike of variance swap to be non-negative. This means the cap multiplier m_{cap} cannot be set too high when the default probability is not small. The intuition is that when the cap level is high, the buyer of the variance swap will be rewarded a decent payoff at the case of default. The issuer of the variance swap would be unable to solve a fair strike to compensate the default-cap payoff when the default probability is not small.

2. The market practice is to set $m_{cap} = 2.5^2$. In our model, such a multiplier is valid only when default probability p is less than 13.8%. Consider a one-year variance swap contract. This corresponds to a CDS spread of around 890 bps with a flat term structure[1]

3. We see that the fair strike is increasing with default probability. This provides an explanation on why the fair strike of variance swap is trading at a premium over its theoretical value using the replication approach, when the probability of default for most companies were not negligible during the recent financial crisis of 2008.

It would be helpful to see how this theoretical result applies on a simple model. For the case of defaultable Black-Scholes model, we have

$$\Psi_x\left(\phi; t_{i-1}, t_i, T\right) = e^{\left(\frac{1}{2}v_0\phi^2 + \left(r_0 + \lambda_0 - \frac{1}{2}v_0\right)\phi\right)\left(t_i - t_{i-1}\right) - \lambda_0 T}, \tag{4.29}$$

such that the pre-default squared-return is given by

$$\rho\left(t_{i-1}, t_i; T\right) = e^{-\lambda_0 T} v_0\left(t_i - t_{i-1}\right).$$

Hence, we can compute the realized variance and adjustment factor γ explicitly as

$$\frac{A}{N}\sum_{i=1}^{N}\rho\left(t_{i-1}, t_i; T\right) = \frac{A}{N}\sum_{i=1}^{N} e^{-\lambda_0 T} v_0\left(t_i - t_{i-1}\right) = v_0 e^{-\lambda_0 T},$$

[1]Take a constant default intensity, we have $p = 1 - e^{-\lambda_0 T}$ and the CDS spread is approximately given by $s = \lambda_0\left(1 - R\right)$, where R is the recovery rate. Plugging in $T = 1$, $p = 1/\left(1 + m\right)$, $m = 2.5^2$ and $R = 40\%$, we obtain $s = 890$ bps.

Stochastic Volatility (X_t^1)	ρ	κ_1	θ_1	σ_1	x_0^1
	-0.3	5	0.12	0.2	0.09
Latent Factor (X_t^2)	κ_2	θ_2	σ_2	x_0^2	
	0.5	0.05	0.2	0.02	
Jump-to-default (λ_t)	λ_0	$\lambda_{1,1}$	$\lambda_{1,2}$		
	0	0	1		

TABLE 4.1: Basic model parameters for the three-factor hybrid model.

and

$$\gamma = \mathbb{E}_x\left[\mathbf{1}_{\{\tau>T\}}\right] - m_{cap}\mathbb{E}_x\left[\mathbf{1}_{\{\tau\leq T\}}\right] = e^{-\lambda_0 T} - m_{cap}\left(1 - e^{-\lambda_0 T}\right),$$

respectively. Therefore, the fair-strike of capped variance swap can be approximated by

$$K = \frac{v_0 e^{-\lambda_0 T}}{e^{-\lambda_0 T} - m_{cap}\left(1 - e^{-\lambda_0 T}\right)} \approx v_0 + m_{cap} v_0 \left(e^{\lambda_0 T} - 1\right). \tag{4.30}$$

It is interesting to see that the fair strike is adjusted by a term $m_{cap}v_0\left(e^{\lambda_0 T} - 1\right)$, which represents the expected payoff from hitting the cap level at $m_{cap}v_0$ with default probability $\left(e^{\lambda_0 T} - 1\right)$.

4.5 Numerical Illustration

This section presents the numerical illustration on the pricing of capped variance swap using the three-factor equity-credit hybrid model. The base model parameters of the three-factor model are given in Table 4.1. For the model parameters, we assume the stochastic volatility has a faster mean-reversion speed (half-life of 0.2 year) than the latent factor of default intensity (half-life of 2 years). For the default intensity, we assume an initial hazard rate of 0.02 which implies an CDS spread of around $100 - 150$ bps. Furthermore, we take the constant interest rate to be 2% and the current stock price to be 100. The time horizon of the variance swap contract is taken to be one-year, in which we consider daily monitoring (252 days), weekly monitoring (52 days) and monthly monitoring (12 days).

4.5.1 Simulation study

First of all, we check our analytical results against the Monte-Carlo simulation by evaluation the accumulation of variance in the presence of jump-to-default. To focus on the impact of default risk, we take the interest rate to be zero. As for the Monte-Carlo simulation, we apply a simple Euler discretization to the three-factor model, with the number of trials is 100,000 and the time-step size kept at 0.001. The estimates reported in Table

Monitoring frequency	Exact	MC Simulation	Std. Errors
Quarterly	36.07	36.06	0.05
Monthly	35.98	35.96	0.03
Weekly	35.94	35.91	0.02
Daily	35.93	35.93	0.02

TABLE 4.2: Analytical and Monte-Carlo esimates of the defaultable realized variance.

4.2 demonstrate that the closed-form solution based on the forward MGF approach is efficient and numerically consistent with the Monte-Carlo simulation.

Next, we investigate the probability of the realized variance of hitting the cap without through its distribution. In particular, we want to verify the assumption

$$\Pr\left\{ I\left(0,T;N\right) > c | \, \tau > T \right\} \approx 0,$$

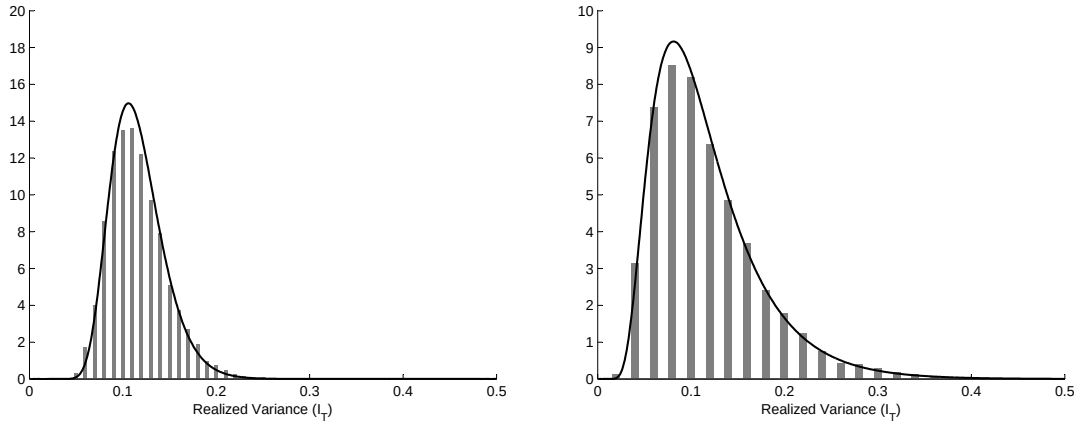
that is, the excess probability that the realized variance is larger than the cap conditional on no default before time T. If this assumption holds well for a wide range of model parameters, then it is effective to compute the fair strike based on the procedure in Section 4.4. Recall that the our stochastic volatility factor is given by:

$$dX_t^1 = \kappa_1 \left(\theta_1 - X_t^1\right) dt + \sigma_1 \sqrt{X_t^1} dW_t^1,$$

we can compute the distribution the continuous-monitoring realized variance based on the fourier inversion of the marginal MGF of I_T. As a benchmark check, we simulate the realized variance with daily-monitoring based on Monte-Carlo technique with 10,000 iterations, time step of 1/10 per day and 252 days in a year. The default event can be simulated by computing the cumulative hazard rate and generating an independent exponential variable as in (4.6). Then, we can compute the marginal distribution of I_T conditional on no default by discarding the defaulted paths, and generate the histogram after normalizing the number of counts with the remaining paths.

From Figure 4.1, we see that the continuous-monitoring realized variance is a good proxy of the discrete realized variance when the monitoring frequency is daily. It is noted that the approximation accuracy deteriorates somewhat when the volatility-of-volatility is small. This is because a smaller volatility-of-volatility implies a slower time-scale of the variance dynamics with respect to the monitoring frequency. Having said that, it appears that the continuous-monitoring is good enough for practical purpose because almost all variance swaps in the market are calculated on daily-monitoring basis.

Figure 4.2 illustrates more clearly the cap-hitting probability $\Pr\left\{ I_T > c | \, \tau > T \right\}$ under continuous-monitoring. As can be seen, the probability of hitting the cap level at 0.2 or



(A) $\sigma_1 = 0.5$. (B) $\sigma_1 = 1.0$.

FIGURE 4.1: The distribution of the realized variance with continuous- and discrete-monitoring.

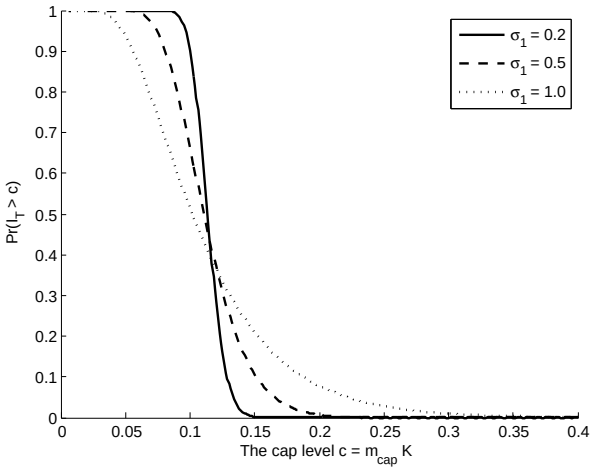


FIGURE 4.2: The excess probability of continuous-monitoring realized variance.

beyond is negligible under reasonable volatility-of-volatility parameters such as $\sigma_1 = 0.2$, 0.5 and 1.0. In particular, these parameters are chosen to satisfy the Feller condition as $2\kappa_1\theta_1 > \sigma_1^2$ which is an important restriction on the square-root volatility model (see Chapter 6 for details). Given that the expected variance is 0.12 (the long-term mean for factor X_t^1), this suggests that the assumption holds well when the multiplier m_{cap} is set to be 1.5^2 or beyond. This guides our analysis on the capped variance swap below, including the impact of the cap multiplier and the pricing behavior of capped variance swap in the presence of interaction between volatility risk and default risk.

4.5.2 The impact of default risk

Figure 4.3 plots the fair-strike of variance swap against the hazard rate of default (λ) with various levels of the multiplier as 1.5^2, 2^2, 2.5^2 and 3^2. In the illustration, when the

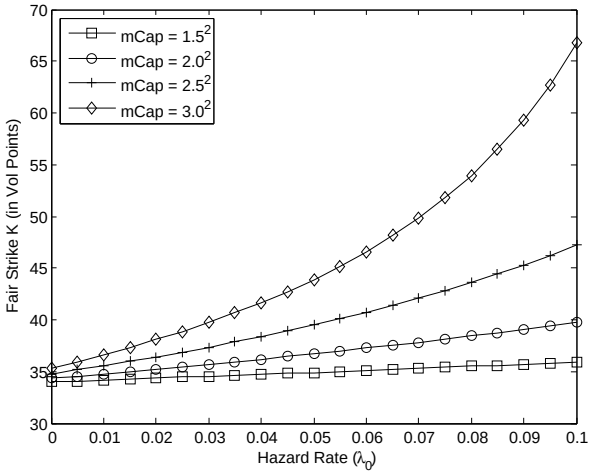


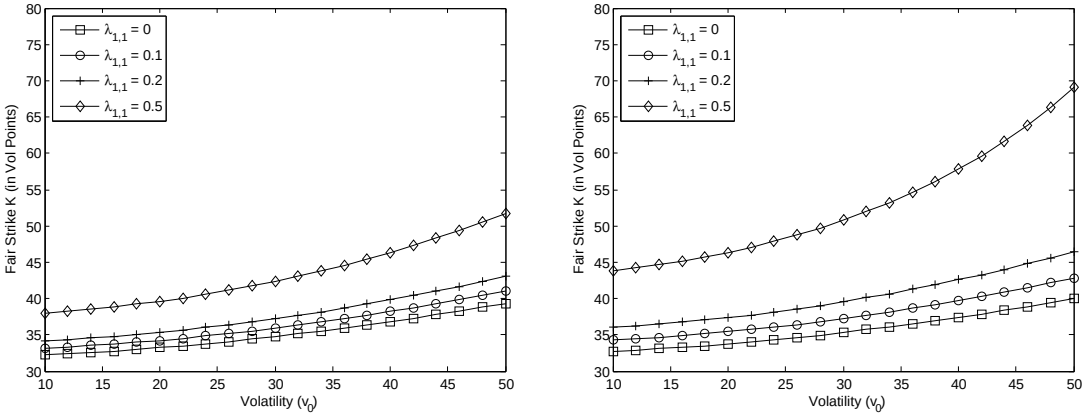
FIGURE 4.3: Fair strike of capped variance swap versus the hazard rate of default.

default intensity λ is set at 2%, the fair strike without default is $\sqrt{K} = 37.7\%$ in terms of volatility points. Suppose we approximate the add-on to the fair-strike in terms of variance by $p \times (m_{cap}K)^2$, we obtain an estimate of 39.2%, which is very close to the theoretical value of 39.1%. The rule-of-thumb of adding a small premium to take into account the cap appears to be a good compromise when the default probability is small.

Let us see what happens when the default cap multiplier is set higher to be 3^2 (i.e., the cap is at $(3K)^2$). In this case, the fair strike is very sensitive to the hazard rate of default. The fair strike increases in a non-linearly manner as the hazard rate goes beyond 5%. This is because the variance swap seller has to set a higher fair strike in order to compensate the risk of a decent payoff at a default event. In contrast, when the multiplier is set lower at 2, the fair strike increases almost linearly. This suggests that the credit risk sensitivity of a capped variance swap depends critically on the choice of the multiplier. In the presence of a high multiplier, a sharp increase in default probability could lead to a significant loss of the variance swap seller. The non-linear sensitivity suggests that one has to be cautious in the basis-risk of constructing a cross-hedging using capped variance swaps and credit default swaps.

4.5.3 Interaction between volatility risk and default risk

Lastly, we illustrate the impact of the interaction between stochastic volatility and hazard rate of default. In particular, we want to examine the sensitivity of the fair strike of capped variance swap with respect to various degrees of volatility-credit interaction. To this end, we take $\lambda_{1,0} = \lambda_{1,2} = 0$, and gradually increase the loading coefficient $\lambda_{1,1}$, in order to generate different levels of interaction. Since we set $\lambda_{1,2} = 0$, our focus is



(A) Cap Multiplier = 2.5^2 (B) Cap Multiplier = 3^2

FIGURE 4.4: Fair strike versus volatility when volatility interacts with hazard rate of default.

on the default risk that is driven purely by the stochastic volatility factor X_t^1 (business risk).

Figure 4.4 plots the fair strike against the instantaneous volatility $X_0^1 = \nu_0$ at different level of volatility-credit interaction as captured by $\lambda_{1,1}$. In general, we find that the impact of the volatility-credit interaction is mild to the pricing of capped variance swap. The impact of endogenous jump-to-default can be quantitative significant only when the volatility-credit interaction is strong and the multiplier is beyond 2.5^2.

4.6 Summary

The joint modeling of equity risk and credit exposure is important in any state-of-the-art option pricing models of credit-sensitive equity derivatives. Our proposed equity-credit models attempt to perform pricing of equity and credit derivatives under a unified framework. We have demonstrated the robustness of adding the jump-to-default feature in the popular affine jump-diffusion models for pricing defaultable European claims and credit default swaps. By assuming the hazard rate to be affine, analytic tractability in typical affine jump-diffusion models is maintained even with the inclusion of the jump-to-default feature. Once the analytic formula is available for the characteristic function of the joint equity-credit price dynamics, numerical valuation of the derivative prices can be performed easily using a standard numerical integration quadrature or Fast Fourier transform algorithm.

For application, we study the pricing of capped variance swap which is standard contractual feature of single-name variance swap trading in the over-the-counter market, in particular after the Lehman crisis. Our numerical example shows that the impact of default

risk on the fair strike of variance swap can be highly non-linear when jump-to-default risk is high. Our results indicate that pricing models that ignore the jump-to-default risk is inadequate for the valuation and risk management of single-name variance swap.

4.7 Appendix

4.7.1 Proof of Lemma 4.2

For $t \in [0,T]$, we have $\mathbf{1}_{\{\tau>T\}} = \mathbf{1}_{\{\tau>t\}}\mathbf{1}_{\{\tau>T\}}$. By the tower property of conditional expectation

$$\begin{aligned}\mathbb{E}_x\left[\pi_t \mathbf{1}_{\{\tau>T\}}\right] &= \mathbb{E}_x\left[\pi_t \mathbf{1}_{\{\tau>t\}}\mathbb{E}_{X_t}\left[\mathbf{1}_{\{\tau>T\}}\middle|\, \mathbf{1}_{\{\tau>t\}}\right]\right] \\ &= \mathbb{E}_x\left[d_{t,T}\pi_t \mathbf{1}_{\{\tau>t\}}\right],\end{aligned}$$

in which $d_{t,T}$ is the forward survival probability from t to T given by

$$d_{t,T} = \mathbb{E}\left[\mathbf{1}_{\{\tau>T\}}\middle|\, X_t, \mathbf{1}_{\{\tau>t\}}\right] = \exp\left[A\left(T-t,\mathbf{0},\mathbf{0}\right) + \left\langle B\left(T-t,\mathbf{0},\mathbf{0}\right), X_t\right\rangle\right],$$

where we apply Proposition 4.1 by noting that X is a homogenous Markov process.

4.7.2 Proof of Theorem 4.3

From the identity

$$\left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 = \frac{\partial^2}{\partial\phi^2}\exp\left[\phi\left(s_{t_i}-s_{t_{i-1}}\right)\right]\Big|_{\phi=0},$$

by Lemma 4.2 and the Fubini's theorem, we have

$$\begin{aligned}\rho\left(t_{i-1},t_i;T\right) &= \mathbb{E}_x\left[\left[\ln\left(S_{t_i}/S_{t_{i-1}}\right)\right]^2 \mathbf{1}_{\{\tau>T\}}\right] \\ &= \mathbb{E}_x\left[\frac{\partial^2}{\partial\phi^2}\exp\left[\phi\left(s_{t_i}-s_{t_{i-1}}\right)\right]\Big|_{\phi=0}\mathbf{1}_{\{\tau>T\}}\right] \\ &= \frac{\partial^2}{\partial\phi^2}\Psi_x\left(\phi;t_{i-1},t_i,T\right)\Big|_{\phi=0}\end{aligned}$$

with

$$\Psi_x\left(\phi;t_{i-1},t_i,T\right) = \mathbb{E}_x\left[\exp\left(A_1 + \left\langle B_1, X_{t_i}\right\rangle\right)\exp\left(\phi\left(s_{t_i}-s_{t_{i-1}}\right)\right)\mathbf{1}_{\{\tau>t_i\}}\right]$$

where $A_1 = A(T-t, \mathbf{0}_n, \mathbf{0}_3)$ and $B_1 = B(T-t, \mathbf{0}_n, \mathbf{0}_3)$ are solution to the Riccati equation in Proposition 4.1.

To compute the expectation, we substitute the definition of the log-stock price process $s_t = \langle \delta_1, X_t \rangle + R_t + \Lambda_t$ for $t = t_i$ and $t = t_{i-1}$ such that

$$\begin{aligned} & \Psi_x(\phi; t_{i-1}, t_i, T) \\ = \; & \mathbb{E}_x\left[e^{\langle B_1+\phi\delta_1, X_{t_i}\rangle + \phi\left(R_{t_i}-R_{t_{i-1}}\right)+\phi\left(\Lambda_{t_i}-\Lambda_{t_{i-1}}\right)} e^{A_1-\phi\delta_1 X_{t_{i-1}}} \mathbf{1}_{\{\tau>t_i\}}\right] \\ = \; & \mathbb{E}_x\left[e^{A_1-\phi\delta_1 X_{t_{i-1}}} \mathbf{1}_{\{\tau>t_{i-1}\}} \mathbb{E}\left[e^{\langle B_1+\phi\delta_1, X_{t_i}\rangle + \phi\left(R_{t_i}-R_{t_{i-1}}\right)+\phi\left(\Lambda_{t_i}-\Lambda_{t_{i-1}}\right)} \mathbf{1}_{\{\tau>t_i\}} \middle| X_{t_{i-1}}, \mathbf{1}_{\{\tau>t_{i-1}\}}\right]\right], \end{aligned}$$

in which we use the tower property of conditional expectation in the second line. Define $\widetilde{R}_t = R_t - R_{t_{i-1}}$ and $\widetilde{\Lambda}_t = \Lambda_t - \Lambda_{t_{i-1}}$ for $t \geq t_{i-1}$, such that X_t and $\widetilde{Y}_t = \left(\widetilde{R}_t, \widetilde{\Lambda}_t\right)$ are joint Markov process and $\widetilde{Y}_{t_{i-1}} = 0$. By the Markov property, we can apply Proposition 4.1 and the inner expectation becomes

$$\mathbb{E}\left[e^{\langle B_1+\phi\delta_1, X_{t_i}\rangle + \phi\left(R_{t_i}-R_{t_{i-1}}\right)+\phi\left(\Lambda_{t_i}-\Lambda_{t_{i-1}}\right)} \mathbf{1}_{\{\tau>t_i\}} \middle| X_{t_{i-1}}, \mathbf{1}_{\{\tau>t_{i-1}\}}\right] = e^{A_2+\langle B_2, X_{t_{i-1}}\rangle},$$

with

$$A_2 = A\left(t_i - t_{i-1}, B_1 + \phi\delta_1, \mathbf{w}_\phi\right), \quad B_2 = B\left(t_i - t_{i-1}, B_1 + \phi\delta_1, \mathbf{w}_\phi\right),$$

with $\mathbf{w}_\phi = (\phi, \phi, 0)$. As a result, we have

$$\Psi_x(\phi; t_{i-1}, t_i, T) = \mathbb{E}_x\left[e^{A_1-\phi\delta_1 X_{t_{i-1}}} e^{A_2+\langle B_2, X_{t_{i-1}}\rangle} \mathbf{1}_{\{\tau>t_{i-1}\}}\right].$$

Finally, by using Proposition 4.1 once again, the expectation can be computed as

$$e^{A_1+A_2}\mathbb{E}_x\left[e^{\langle B_2-\phi\delta_1, X_{t_{i-1}}\rangle} \mathbf{1}_{\{\tau>t_{i-1}\}}\right] = e^{A_1+A_2+A_3} e^{\langle B_3, x\rangle},$$

with

$$A_3 = A\left(t_{i-1}, B_2 - \phi\delta_1, \mathbf{0}_3\right), \quad B_3 = B\left(t_{i-1}, B_2 - \phi\delta_1, \mathbf{0}_3\right).$$

Chapter 5

Pricing Models of Contingent Convertibles

5.1 Introduction

The contingent convertible bond (CoCo) is a hybrid fixed income security that provides a loss-absorption mechanism when the capital of the issuing bank falls close to the regulatory level as required by the Basel Committee on Banking Supervision (BCBS). At a triggering event, the bond is automatically converted into the equity of the issuing bank (or an equivalent amount of cash). The conversion provides fresh capital to the issuing bank and saves it from financial distress. As a result, this can help to mitigate the chance of a systemic banking crisis while avoiding the use of taxpayer's money to bail out distressed financial institutions.

Since the first issuance of Enhanced Capital Notes by the Lloyds Banking Group in December 2009, there has been active discussion on the triggering mechanism and loss-absorption design of a CoCo bond. There are three possible triggering mechanisms of a CoCo bond: accounting trigger, market trigger and regulatory trigger. In an accounting trigger, the capital ratio is chosen as the indicator on a bank's financial health. For example, when the tier-1 capital ratio falls below a certain level, the bond is converted into the equity in order to boost the bank's capitalization. In a market trigger, the stock price is proposed as a forward-looking indicator on the financial health of a bank. When the stock price falls below a pre-defined barrier level, the CoCo bond is triggered automatically. Finally, in a regulatory trigger, sometimes called the point of non-viability (PONV) trigger, the banking supervisory authority holds the discretion to judge whether a bank is insolvent or not, and determine accordingly whether a conversion should be activated. Having said that, it appears that the financial market has chosen the

accounting trigger as the *de-facto* mechanism as noted in a recent survey by Avdjiev et al. (2013).

There are a number of papers that address the pricing and risk management of a CoCo bond using various modeling approaches. The structural approach starts with the modeling of a bank's balance sheet dynamics which allows one to analyze the impact of the issuance of contingent convertibles on the capital structure (Albul et al., 2010; Brigo et al., 2013; Glasserman and Nouri, 2012; Pennacchi, 2011). Cheridito and Xu (2014) apply the reduced-form approach, commonly used for the pricing of credit derivatives, to the evaluation of CoCo bonds. The reduced-form approach has the great flexibility to match the market prices of CoCo bonds and can be extended to perform calibration with the credit default swap (CDS) spreads. Meanwhile, Spiegeleer and Schoutens (2012) propose an easy-to-use equity derivative approach by approximating the accounting trigger of the capital ratio by the first-passage-time of the stock price process to an implied barrier level. Along this line, Corcuera et al. (2013) pursue a smile conforming model that assumes the stock price follows a Levy process. Gupta et al. (2013) discuss the treatment on various contractual features of CoCo bonds and resort to numerical methods for the pricing of CoCo bonds with a mean-reverting capital ratio. Wilkens and Bethke (2014) report the empirical assessment of aforementioned approaches and find that the equity derivative approach implies a hedging ratio that is practically useful, in view of the risk management of CoCo bonds during the sample period of 2011.

The modeling of the capital ratio is essential to respect the contractual design of a CoCo bond with accounting ratio trigger. Hence, the structural modeling approach provides a natural starting point to the pricing of CoCos since the capital ratio is after all a balance sheet quantity. The structural approach, however, does not usually possess the flexibility for the calibration to traded security prices and fails to generate a reasonable shape of the credit spread. Furthermore, as equity is priced as a contingent claim, the joint dynamics of the stock price and capital ratio is not tractable under the structural framework. It is also not straight-forward to incorporate a jump in the stock price, which further restricts its ability to reflect the potential write-down of the CoCo bond value upon a conversion. On the other hand, the reduced-form approach is efficient for the pricing of CoCos and it only requires one to specify the conversion intensity and the size of the stock price jump at the conversion time. The approach is expected to work well when the capital ratio is far away from the triggering threshold. In this case, the CoCo bond is mainly consisted of a straight bond component while the equity component is only a residual. As such, the conversion can be treated as a rare event which can be approximated by a single Poisson jump. A criticism on the reduced-form approach is that it completely ignores the contractual feature of an accounting trigger and is silent about the interaction in between stock price and capital ratio.

In this chapter, we propose a bivariate modeling of the stock price and the capital ratio which is enhanced by a jump-to-non-viability (JtNV) feature. We follow the notion in Kijima et al. (2009a) in which the stock price is interpreted as a marker process which is correlated to the capital ratio. Such a joint modeling is particularly important to the pricing of CoCos because the conversion value depends on both the first-passage-time of the capital ratio towards the triggering threshold and the stock price at the conversion. Since we directly model the two most important quantities for the pricing of CoCos, the parametrization is much easier than a bottom-up structural model which requires a judious choice of the proxy conversion epoch as linked to other balance sheet quantities (Brigo et al., 2013; Pennacchi, 2011). We demonstrate that the equity derivative approach proposed by Spiegeleer and Schoutens (2012) can be recast under our proposed bivariate framework when the stock price and the capital ratio are perfectly correlated. This provides the justification to the equity derivative approach that replicates of CoCo components using barrier options. Furthermore, we add in the reduced-form feature that is important to capture a JtNV which can be related to a sudden insolvency of the bank leading to a trigger. In the aftermath of the Lehman financial crisis, it is not uncommon to see substantial write downs by major investment banks due to unexpected trading losses and bleaching of regulation, which might erode a significant part of the bank's capital. In summary, the proposed approach that integrates the reduced-form approach and structural approach is natural for the pricing of a CoCo bond since it is a hybrid financial product that has multiple sources of risk.

We remark that the proposed framework is an extension of the equity-credit hybrid modeling framework in Chapter 4 which allows one to model the interaction between equity risk and credit risk (Carr and Linetsky, 2006; Carr and Wu, 2009; Cheridito and Wugalter, 2012; Chung and Kwok, 2014). For the sake of practical implementation, we illustrate models that are easy-to-implement yet flexible enough to calibrate with the market prices of CoCos. When the capital ratio follows a simple lognormal process, we manage to derive the closed-form pricing formula which can be seen as an extension to Black and Cox (1976) on a new class of corporate security. When the capital ratio is mean-reverting, we can express the pricing formula in terms of the Laplace transform based on standard results on the first-passage-time density of an OU process. In addition, we demonstrate the usage of a simple numerical algorithm known as the Fortet method, which has been employed in Longstaff and Schwartz (1995), Collin-Dufresne and Goldstein (2001) and Coculescu et al. (2008). While it is possible to extend to more sophisticated dynamics of stock price and capital ratio such as those with stochastic volatility and jump diffusion, the infrequent observation and the lack of historical data of the capital ratio make it difficult to justify the pursue of these advanced models.

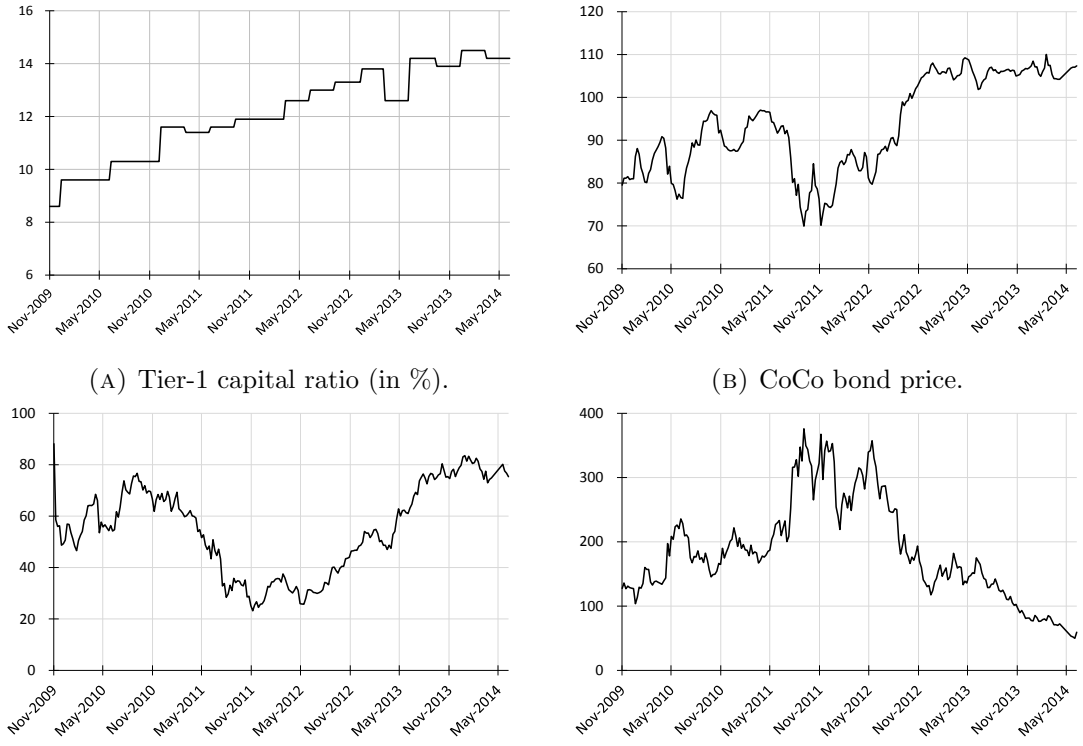
5.2 Pricing of a CoCo Bond

5.2.1 Recent development

In this section, we take the Lloyds banking group (LBG) as an example because its issued CoCo bonds have the longest historical data and have been actively traded in the market. Figure 5.1 reports the historical time-series of the tier-1 capital ratio, CoCo bond price, stock price and 5-year CDS spread for the LBG. The CoCo bond is the LBG 7.8673% enhanced capital note with the maturity on 17 December 2019 and a tier-1 capital ratio trigger at 5% (ISIN: XS0459086749). Since its first issuance, the bank's tier-1 capital ratio has been gradually increasing and staying at a healthy level away from the 5% triggering level, with the CoCo bond price increased by more than 30%. Interestingly, the capital ratio does not co-move perfectly with the stock price and they might diverge from time-to-time. The pairwise correlation in between the stock price and the capital ratio is estimated to be 0.43 (using quarterly log-returns from 2009Q1 to 2014Q2). Hence, the correlation is less than perfect and this does not support the notion of replacing an accounting trigger by a stock price trigger in the equity derivative approach.

Meanwhile, we find that the CoCo bond price depicts strong co-movements with the stock price and the CDS spread, with the pairwise correlations estimated to be 0.51 and -0.70 respectively (using quarterly log-returns from 2009Q1 to 2014Q2). This is not surprising given the hybrid nature of a CoCo bond with exposure to both equity risk and default risk. When the CDS spread escalated to the level of 300 bps during August 2011 as the European debt crisis unfolded, we see that both the CoCo bond price and stock price tumbled together although the capital ratio remained strong at a level of 12%. This suggests that market participants might have anticipated the possibility of a sudden insolvency of the bank that could trigger the conversion of the CoCo bond during these distress periods albeit the apparently strong capital adequacy. Moreover, the empirical relationship of the CoCo bond price with the CDS spread suggests that it might be effective to hedge a CoCo bond using credit instruments.

Since late 2012, the CoCo bond price appears to move into a fixed income regime close to its bond ceiling: the capital ratio is distant from the trigger (at a level of 14%) and the equity nature of the CoCo bonds is largely ignored in which a conversion being considered as a rare event. It is reportedly that market participants have perceived a CoCo bond primarily as a straight bond paying juicy coupons with a small probability of a large write-down due to a conversion. Most of the issued CoCo bonds have been invested by fixed income fund managers who are seeking for high yields rather than equity exposure. It is expected that the equity component will become more important when the capital



(A) Tier-1 capital ratio (in %).

(B) CoCo bond price.

(C) Stock price (in GBP pence).

(D) CDS spread (in bps).

FIGURE 5.1: Historical time-series of tier-1 capital ratio, CoCo price, stock price and CDS spread for the Lloyds Banking Group. All data are from Bloomberg.

ratio moves closer to the triggering level in the future. In that scenario, an appropriate modeling on the joint dynamics of the capital ratio and stock price becomes essential.

In summary, the CoCo bond price shows strong empirical relationship to the capital ratio, stock price and CDS spreads, and its risk exposure appears to change overtime when the capital ratio is at different levels. This motivates the proposal of an unified framework to incorporate these empirical features.

5.2.2 The structure

The typical structure of a CoCo bond can be described as follows:

1. Bond component: coupon payments $(c_i)_{i=1,2,\ldots n}$ and principal payment F at the maturity T.
2. Conversion component: at a trigger event, the bond is converted into either shares of equity or cash.

In an equity conversion, the bond is converted into G shares of equity and this introduces the equity exposure to the CoCo bond. For a cash conversion, the bond converts into G

units of the currency and hence there is no direct equity exposure. In general, the pricing of CoCos is related to interest rate risk (due to the coupon and principal payments), equity risk (if the bond is converted to equity at the trigger event) and conversion/default risk. The conversion risk can be interpreted as the risk of an unfavorable conversion to a declined stock price that wipes off the value of the bond. We emphasize that the pricing of CoCos is *indirectly* related to default risk because a conversion always happens before a bank's default. In particular, we only need the knowledge of the stock price up to the conversion time for the pricing of CoCos.

Denote the conversion time to be τ and the stock price process to be $S = (S_t)_{t\geq 0}$, and assume a constant interest rate r. Under the risk-neutral measure Q, the no-arbitrage price of a CoCo can be decomposed as

$$P_{CoCo} = P_C + P_F + P_E$$

with following components:

1. Coupon payment P_C:

$$P_C = \sum_{i=1}^{n} c_i \mathbb{E}^Q \left[e^{-rt_i} \mathbf{1}_{\{\tau > t_i\}} \right] = \sum_{i=1}^{n} c_i e^{-rt_i} \left[1 - Q\left(\tau \leq t_i\right) \right],$$

   which is the sum of coupon payments before the conversion time τ.[1]

2. Face value P_F :

$$P_F = F\mathbb{E}^Q \left[e^{-rT} \mathbf{1}_{\{\tau > T\}} \right] = F e^{-rT} \left[1 - Q\left(\tau \leq T\right) \right],$$

   in which F is the principal payment when there is no conversion until maturity.

3. Conversion value P_E : at the conversion, the CoCo is converted into G-shares of the underlying equity

$$P_E = G\mathbb{E}^Q \left[e^{-r\tau} S_\tau \mathbf{1}_{\{\tau \leq T\}} \right].$$

Therefore, we need to compute the conversion probability to evaluate the bond component consisted of the coupon payments and face value, while resorting to the joint density of the stock price and the conversion time in order to compute the conversion value P_E. Hence, the key to the pricing of a CoCo bond is the joint modeling of the conversion time τ and stock price S_t. For the conversion-into-cash, we only need to replace the term GS_τ by a constant cash payment G and this is an easier problem.

[1] Note that we ignore the interest accrual when a conversion happens in between two coupon payment dates which is expected to be negligible.

5.3 An Enhanced Hybrid Modeling

5.3.1 Setup

We take the stock price to be $S_t = \exp(x_t)$ and capital ratio to be $H_t = \exp(y_t)$, where $X_t = (x_t, y_t)$ follows a bivariate process under the risk-neutral measure Q as

$$\begin{aligned} dx_t &= \left(r - q - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t^1 + \gamma \left(dN_t - \lambda_t dt\right), \qquad x_0 = x, \\ dy_t &= \kappa\left(\theta - y_t\right) dt + \eta\left(\rho dW_t^1 + \sqrt{1-\rho^2} dW_t^2\right), \qquad y_0 = y, \end{aligned} \tag{5.1}$$

in which W_t^1 and W_t^2 are independent Brownian motions. The stock price is a jump diffusion process with the constant jump size γ and the intensity λ_t in order to capture the write-down in the stock price upon a sudden insolvency. The capital ratio is modelled as an exponential Ornstein-Unlenbeck (OU) process which naturally incorporates mean reversion. This is because banks usually actively manage the amount of regulatory capital in response to the changing market values of asset and liability, and they have the incentives to maintain a healthy level of capital ratio in order to avoid any regulatory bleach.

Following the notion in Kijima et al. (2009a) and Siu and Kijima (2014), the stock price process x_t is taken to be an observable marker process which is correlated to the latent capital ratio process y_t. To illustrate the idea, let us take the equivalent parameters by setting $\overline{\eta} = \rho\eta$ and $\varepsilon = \eta\sqrt{1-\rho^2}$. The dynamics of the latent process y_t can be expressed as

$$dy_t = \kappa\left(\theta - y_t\right) dt + \overline{\eta} dW_t^1 + \varepsilon dW_t^2,$$

in which the second diffusion term εdW_t^2 can be interpreted as the noise due to the latent feature of y_t. When $\rho = 1$, we have $\varepsilon = 0$ and the dynamics of y_t is known once we know the path of x_t. When ρ decreases from 1 to zero, we have a higher degree of noisy information about the latent factor y_t. We can perform the pricing of financial instruments since the joint law of the bivariate process (x_t, y_t) under the risk-neutral measure Q is well-specified by (5.1).[2]

The accounting trigger is defined as the first-passage-time of the log capital ratio y_t to a lower threshold y_B as

$$\tau_B = \inf\left\{t \geq 0;\, y_t = y_B\right\},$$

[2]Another notion of latent process can be found in Duffie and Lando (2001) and Coculescu et al. (2008) which involve a more sophisticated characterization of the information structure available to the investors. Nevertheless, these models are less tractable in the pricing of complicated financial instruments and we left this for future research.

where $H_B = \exp(y_B)$ is the contractual threshold for an accounting trigger. On the other hand, we introduce the JtNV feature by using the first jump of the Poisson process N_t with the intensity parameter $\lambda_t = \lambda(X_t)$ as

$$\tau_R = \inf\{t \geq 0;\, \Lambda_t \geq e\}, \qquad \Lambda_t = \int_0^t \lambda(X_u)\, du,$$

in which e is an independent standard exponential variable. Hence, the conversion time is taken to be the earlier of the first-passage-time τ_B or the JtNV time τ_R as

$$\tau = \tau_B \wedge \tau_R.$$

Without loss in generality, we assume that $\Pr(\tau_B = \tau_R) = 0$, i.e., the two random times are different almost surely.

It is worth to make a remark on the change in stock price upon a conversion. At the accounting trigger $\tau = \tau_B$, the stock price is continuous so that $S_\tau = S_{\tau^-}$. The rationale is that the stock price should have gradually taken into account the possibility of such a conversion. At the regulatory trigger $\tau = \tau_R$, there is a fixed jump in stock price as $S_\tau = (1+\gamma) S_{\tau^-}$, where $\gamma \in (-1, \infty)$. Hence, we have the following three cases:

1. Jump-to-default when $\gamma = -1$;
2. Jump-to-partial-default when $-1 < \gamma \leq 0$;
3. Jump-to-recover when $\gamma > 0$.

In practice, it is not easy to determine the parameter γ because it is not sure whether the sudden conversion is a good news or bad news to the equity holders. A conversion reduces the liability of the bank by canceling the coupon and principal payments of the bond, which in turn boosts the capital adequacy and leads to a stronger balance sheet and stock price. On the other hand, the conversion into shares of stock causes dilution to existing equity holders which should be reflected by a weakened stock price. The actual effect of the conversion will only become clear when there is an actual conversion event in the future.[3] Lastly, note that we have made no assumption about the change of the capital ratio after the JtNV as this is not related to the pricing.

The next lemma gives the risk-neutral conversion probability $Q(\tau \leq t)$ under the enhanced hybrid modeling framework. The complicated forms of the formula results from the convolution of the two stopping times τ_B and τ_R.

[3]We note that it is not meaningful to take $\gamma = -1$, i.e., jump-to-default. This is because (i) that implies the regulatory-triggered conversion completely wipes off the value of equity, and (ii) the CoCo conversion is designed to absorb loss and prevent default of the bank.

Lemma 5.1. *For a fixed $t \geq 0$, the conversion probability is given by*

$$Q(\tau \leq t) = \int_0^t \mathbb{E}^Q \left[\lambda(X_u) e^{-\int_0^u \lambda(X_s) ds} \mathbf{1}_{\{\tau_B > u\}} \right] du + \mathbb{E}^Q \left[e^{-\int_0^{\tau_B} \lambda(X_u) du} \mathbf{1}_{\{\tau_B \leq t\}} \right].$$

Proof. See Appendix. □

From Lemma 5.1, it is noted that

1. When there is no JtNV trigger by setting $\lambda(X_t) = 0$, we have

   $$Q(\tau \leq t) = \mathbb{E}^Q \left[\mathbf{1}_{\{\tau_B \leq t\}} \right],$$

   and the conversion is only due to an accounting trigger.

2. When the capital ratio is far away from the triggering threshold or the intensity $\lambda(\cdot)$ is sufficiently large, we can approximate

   $$\begin{aligned} Q(\tau \leq t) &\approx \int_0^t \mathbb{E}^Q \left[\lambda(X_u) e^{-\int_0^u \lambda(X_s) ds} \right] du \\ &= \mathbb{E}^Q \left[\int_0^t \lambda(X_u) e^{-\int_0^u \lambda(X_s) ds} du \right] = \mathbb{E}^Q \left[\mathbf{1}_{\{\tau_R \leq t\}} \right], \end{aligned}$$

   and the conversion is only due to a JtNV jump. Moreover, this is rendered to a reduced-form model.

3. When the intensity is constant, the conversion probability can be expressed as

   $$\begin{aligned} Q(\tau \leq t) &= \int_0^t \lambda e^{-\lambda u} \left[1 - Q(\tau_B \leq u) \right] du + \mathbb{E}^Q \left[e^{-\lambda \tau_B} \mathbf{1}_{\{\tau_B \leq t\}} \right] \\ &= 1 - e^{-\lambda t} Q(\tau_B > t), \end{aligned}$$

   in which the second equality follows from an integration-by-part.

5.3.2 Conversion value

To explore the analytical tractability of the proposed framework, let us further assume that the JtNV intensity λ is a constant. The key to the CoCo pricing problem is the computation of the conversion value as

$$P_E = G\mathbb{E}^Q \left[e^{-r\tau} S_\tau \mathbf{1}_{\{\tau \leq T\}} \right], \qquad \tau = \tau_B \wedge \tau_R,$$

that is, the expected value when the CoCo is converted into shares of the underlying equity. It is not straight-forward to compute this expectation because the stock price

may experience a jump at $\tau = \tau_B \wedge \tau_R$ due to the JtNV trigger. To this end, it is useful to apply a change-of-measure that simplifies the pricing formula by removing the dependence on the stock price S (Cheridito and Xu, 2014). Define the adjusted stock price process as

$$\begin{aligned}\widetilde{S}_t &= S_t e^{-(r-q)t} \\ &= S_0 \exp\left(\int_0^t \sigma dW_s^1 - \int_0^t \frac{1}{2}\sigma^2 ds - \gamma\Lambda_t\right)(1+\gamma)^{N_t}, \quad t \geq 0.\end{aligned}$$

It is readily to see that $\widetilde{S} = \left(\widetilde{S}_t\right)_{t\geq 0}$ is a positive martingale under the risk-neutral measure Q. Hence, we can take $\widetilde{S}_t$ to construct a change-of-measure as

$$Z_t = \left.\frac{dQ^*}{dQ}\right|_{\mathcal{F}_t} = \frac{\widetilde{S}_t}{\widetilde{S}_0} = \frac{e^{-(r-q)t}S_t}{S_0},$$

where Q^* can be interpreted as the *stock price measure*.

Lemma 5.2. *Under the stock price measure Q^*, the bivariate process evolves as*

$$\begin{aligned}dx_t &= \left(r - q - \frac{1}{2}\sigma^2\right)dt + \sigma dB_t^1 + \gamma\left(dN_t - \lambda dt\right), \qquad x_0 = x, \\ dy_t &= \left(\kappa\left(\theta - y_t\right) + \rho\sigma\eta\right)dt + \eta dB_t^2, \quad y_0 = y,\end{aligned}$$

where $\left\langle dB_t^1, dB_t^2\right\rangle = \rho dt$. The adjusted conversion intensity is given by

$$\lambda^* = (1+\gamma)\lambda.$$

Proof. See Appendix. □

Now, we can apply the change-of-measure to evaluate the conversion value as

$$\mathbb{E}^Q\left[e^{-r\tau}S_\tau \mathbf{1}_{\{\tau\leq T\}}\right] = S_0\mathbb{E}^Q\left[e^{-q\tau}\frac{e^{-(r-q)\tau}S_\tau}{S_0}\mathbf{1}_{\{\tau\leq T\}}\right] = S_0\mathbb{E}^{Q^*}\left[e^{-q\tau}\mathbf{1}_{\{\tau\leq T\}}\right],$$

with $\tau = \tau_B \wedge \tau_R$. As a result, we only need to compute a discounted conversion probability under the stock price measure Q^*. The next two propositions provide the formula to compute the relevant expectation.

Proposition 5.3. *Denote $\lambda^* = (1+\gamma)\lambda$. The conversion value is given by*

$$P_E = GS_0\left\{\int_0^T \lambda^* e^{-(\lambda^*+q)u}\left(1 - Q^*\left(\tau_B \leq u\right)\right)du + \mathbb{E}^{Q^*}\left[e^{-(\lambda^*+q)\tau_B}\mathbf{1}_{\{\tau_B\leq T\}}\right]\right\}.$$

Proof. By Lemma 5.2, we only need to replace λ by $\lambda^* = (1+\gamma)\lambda$ and take the expectation under the stock price measure Q^*. The rest is similar to Lemma 5.1. □

By integration-by-part, we can obtain an alternative representation of the conversion value which is easier to compute. A similar exercise has been performed in Campi et al. (2009) on the equity-credit hybrid modeling using a CEV process.

Proposition 5.4. *When $\lambda^* + q > 0$, the conversion value can be expressed as*

$$P_E = GS_0 \left\{ \frac{\lambda^*}{\lambda^* + q} \left(1 - e^{-(\lambda^*+q)T} Q^* \left(\tau_B > T \right) \right) + \frac{q}{\lambda^* + q} \mathbb{E}^{Q^*} \left[e^{-(\lambda^*+q)\tau_B} \mathbf{1}_{\{\tau_B \leq T\}} \right] \right\}.$$

Proof. See Appendix. □

By Propositions 5.3 and 5.4, we have effectively reduce a two-dimensional problem into a one-dimensional problem using the change-of-measure formula. The key to the pricing of CoCo bond is the evaluation of the first-passage-time probability $Q^* \left(\tau_B \leq t \right)$ and the associated truncated Laplace transform

$$I = \mathbb{E}^{Q^*} \left[e^{-(\lambda^*+q)\tau_B} \mathbf{1}_{\{\tau_B \leq T\}} \right]. \tag{5.2}$$

In the next section, we discuss several specific models that are practically implementable.

5.4 Examples

5.4.1 A simple reduced-form model

First-of-all, let us briefly review the reduced-form approach which is a nested case of the enhanced hybrid approach. Suppose that the conversion intensity is constant and assume zero dividend ($q = 0$), we have a simple pricing formula

$$P_{CoCo} = \sum_{i=1}^{n} c_i e^{-(r+\lambda)t_i} + F e^{-(r+\lambda)T} + GS_0 \left(1 - e^{-(1+\gamma)\lambda T} \right), \tag{5.3}$$

which resembles the pricing formula of a defaultable bond (first two terms) with a recovery payment (the last term). Given the contractual setup of the CoCo bond (c_i, F and G), the remaining model parameters are the conversion intensity λ and the jump size γ for the write-down of stock price upon conversion. The intensity parameter λ has the conventional meaning of credit spread which takes into account the cancellation of coupons and principal payment upon conversion. The jump size γ determines the size of the recovery payment which is the expected value of the stock price at a conversion.

As for the consistent pricing of a CoCo bond, it is useful to calibrate the model parameters using financial instruments in other asset classes, such as CDS spreads and deep

out-of-the-money (DOOM) put options. For instance, one can estimate the implied default intensity from the CDS spread using the rule-of-thumb as

$$\lambda_{CDS} = \frac{c}{1-R}$$

where c is the market CDS spread and R is the recovery rate usually set to be 40%. As the CoCo conversion always occurs before the default of a company (as loss absorption mechanism), we can interpret the CDS implied intensity as a *lower bound* of the CoCo conversion intensity. Alternatively, one might estimate the default intensity from the DOOM put option using the defaultable put option price formula:

$$p(S,\tau) = Ke^{-(r+\lambda)\tau}N(-d_2) - SN(-d_1), \quad \tau = T-t,$$

where K is the strike price, and

$$d_1 = \frac{\ln(S/K) + \left(r+\lambda+\frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau}.$$

Note that the defaultable put option formula is simply the standard Black-Scholes formula by replacing the riskfree interest rate r by the risk-adjusted discount rate $r+\lambda$.

5.4.2 Brownian capital ratio

When $\kappa = 0$, we have the Brownian capital ratio model. It would be more convenient to write down the pre-conversion dynamics of the bivariate process (x_t, y_t) as

$$\begin{aligned} dx_t &= \beta dt + \sigma dW_t^1, \quad x_0 = x, \\ dy_t &= \alpha dt + \eta\left(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2\right), \quad y_0 = y. \end{aligned}$$

with $\alpha = 0$ and $\beta = r - q - \frac{1}{2}\sigma^2 - \lambda\gamma$, so that we can recast it as the bivariate model in Kijima et al. (2009a). Before the conversion event, the stock price evolves as a lognormal process following the Black-Scholes model while the log capital ratio follows a Brownian motion in order to avoid an exponential growth. The Brownian model is highly analytically tractable and leads to simple pricing formula.

Perfect correlation

Take $\rho = 1$, the capital ratio process is perfectly correlated to the stock price process. In this case, the event "capital ratio hitting a lower threshold" is equivalent to the event

"the stock prices hitting a down barrier".[4] Denote the first-passage-times of the capital ratio process and stock price process as

$$\tau_B = \inf\{t \geq 0;\, y_t = y_B\} \quad \text{and} \quad \tau_S = \inf\{t \geq 0;\, x_t = x_B\}$$

respectively, where $y_B = \ln H_B$ and $x_B = \ln S_B$. It can be shown that the two stopping times τ_B and τ_S are equal in distribution for a certain implied threshold $x_B = \ln S_B$.

Furthermore, suppose that we switch off the JtNV feature by setting $\lambda = 0$. We can proxy the conversion probability by the first-passage-time distribution of stock price to the implied threshold $S_B = \exp(x_B)$ as

$$\begin{aligned} Q(\tau_S \leq t) &= N\left[\frac{x_B - x_0 - \beta t}{\sigma\sqrt{t}}\right] \\ &\quad + \exp\left[\left(\frac{2\beta}{\sigma^2}\right)(x_B - x_0)\right] N\left[\frac{x_B - x_0 + \beta t}{\sigma\sqrt{t}}\right], \end{aligned}$$

where $N(\cdot)$ is the standard cumulative normal distribution. By appropriate rearrangement, we can recover the pricing formula under the equity derivative approach as proposed by Spiegeleer and Schoutens (2012).

Imperfect correlation

When $-1 < \rho < 1$, one cannot fully proxy the dynamics of the capital ratio by using the stock price process. We need the knowledge of the first-passage-time τ_B defined as

$$\tau_B = \inf\{t \geq 0;\, y_t = y_B\}.$$

We have the following useful quantities:

1. The density for τ_B:

$$Q\{\tau_B \in dt\} = g(t; y_B)\, dt = \frac{y_0 - y_B}{\sqrt{2\pi\eta^2 t^3}} \exp\left[-\frac{(y_B - y_0 - \alpha t)^2}{2\eta^2 t}\right] dt, \tag{5.4}$$

   for $y_0 > y_B$ (i.e., down-and-out passage time).

2. The distribution for τ_B:

$$Q(\tau_B \leq t) = N\left[\frac{y_B - y_0 - \alpha t}{\eta\sqrt{t}}\right] + \exp\left[\left(\frac{2\alpha}{\eta^2}\right)(y_B - y_0)\right] N\left[\frac{y_B - y_0 + \alpha t}{\eta\sqrt{t}}\right]. \tag{5.5}$$

[4]The other case of perfect correlation with $\rho = -1$ would implies an equivalent up-and-out hitting time for the stock price which is not economically meaningful.

The derivation can be found in the Appendix. The corresponding density and distribution under the stock price measure Q^* are readily obtained by replacing α by $\alpha + \rho\eta\sigma$. The truncated Laplace transform in (5.2) related to the calculation of the conversion value can be obtained in closed-form.

Proposition 5.5. *Take $\zeta = \lambda^* + q > 0$, the truncated Laplace transform in (5.2) can be obtained in closed-form as*

$$\begin{aligned} I &= \mathbb{E}^{Q^*}\left[e^{-\zeta\tau_B}\mathbf{1}_{\{\tau_B\leq T\}}\right] \\ &= e^{\delta^-(y_B-y_0)}N\left[\frac{y_B-y_0-\sqrt{\tilde{\alpha}^2+2\zeta\eta^2}t}{\eta\sqrt{t}}\right]+e^{\delta^+(y_B-y_0)}N\left[\frac{y_B-y_0+\sqrt{\tilde{\alpha}^2+2\zeta\eta^2}t}{\eta\sqrt{t}}\right], \end{aligned}$$

where

$$\tilde{\alpha} = \alpha + \rho\sigma\eta, \quad \textit{and} \quad \delta^{\pm} = \frac{\tilde{\alpha} \pm \sqrt{\tilde{\alpha}^2+2\zeta\eta^2}}{\eta^2}.$$

Proof. See Appendix. □

Remark 5.6. We note that the dividend q enters into the pricing formula in a more subtle way than that in Spiegeleer and Schoutens (2012) who derive the pricing formula based on a replication portfolio using barrier options.

When $\lambda = 0$ and $q = 0$, the conversion value can be expressed as

$$P_E = GS_0Q^*\left(\tau_B \leq T\right),$$

where

$$\begin{aligned} Q^*\left(\tau_B \leq T\right) &= N\left[\frac{y_B-y_0-\left(\alpha+\rho\sigma\eta\right)t}{\eta\sqrt{t}}\right] \\ &\quad + \exp\left(\frac{2(\alpha+\rho\sigma\eta)}{\eta^2}\left(y_B-y_0\right)\right)N\left[\frac{y_B-y_0+\left(\alpha+\rho\sigma\eta\right)t}{\eta\sqrt{t}}\right], \end{aligned} \quad (5.6)$$

is the first-passage-time distribution under the stock price measure Q^*. In this case, we have a simple closed-form CoCo pricing formula

$$P_{CoCo} = \sum_{i=1}^{n} c_i e^{-rt_i}\left[1-Q\left(\tau_B \leq t_i\right)\right] + Fe^{-rT}\left[1-Q\left(\tau_B \leq T\right)\right] + GS_0Q^*\left(\tau_B \leq T\right), \quad (5.7)$$

where $Q\left(\tau_B \leq t_i\right)$ and $Q^*\left(\tau_B \leq T\right)$ are given by Eqs. (5.5) and (5.6) respectively. When the current stock price S_0 is known, the conversion value

$$GS_0Q^*\left(\tau_B \leq T\right)$$

depends on the conversion probability under the stock price measure Q^* in which the drift of the capital-ratio dynamics is adjusted by $\rho\sigma\eta$. In particular, when $\rho = 0$ the conversion probability is the same under Q and Q^*. This means that the conversion value is independent to the future stock price dynamics when the correlation is zero.

5.4.3 Mean-reverting capital ratio

The pricing of a CoCo bond can be performed analytically using the standard results on the first-passage-time of an OU process. Recall that the marginal dynamics of the capital ratio follows the exponential OU process as

$$dy_t = \kappa\left(\theta - y_t\right)dt + \eta d\tilde{W}_t^2, \quad y_0 = y.$$

Denote the infinitesimal generator of the OU process as

$$\mathcal{L} = \frac{1}{2}\eta^2\frac{d^2}{dy^2} + \kappa\left(\theta - y\right)\frac{d}{dy}.$$

Let F and G be the two fundamental solutions of $\mathcal{L}u(y) = \alpha u(y)$ for $y \in \mathbb{R}$ such that (i) F is a positive increasing function and (ii) G is a positive decreasing function. It is known that (see, Borodin and Salminen, 2002)

$$F(y) = F(y;\alpha) = \int_0^\infty u^{\alpha/\kappa - 1} e^{\sqrt{\frac{2\kappa(y-\theta)}{\eta^2} - \alpha^2/2}} du,$$
$$G(y) = G(y;\alpha) = \int_0^\infty u^{\alpha/\kappa - 1} e^{\sqrt{\frac{2\kappa(\theta-y)}{\eta^2} - \alpha^2/2}} du.$$

The Laplace transform of the first-passage-time $\tau_B = \inf\left(t \geq 0;\, y_t = y_B\right)$ under the Q-measure is

$$\mathbb{E}^Q\left[e^{-\alpha\tau_B}\right] = \begin{cases} \dfrac{F(y)}{F(y_B)}, & \text{when} \quad y \leq y_B, \\ \dfrac{G(y)}{G(y_B)}, & \text{when} \quad y \geq y_B. \end{cases}$$

As a result, we can express (Davydov and Linetsky, 2001)

$$\begin{aligned} \mathbb{E}^Q\left[\mathbf{1}_{\{\tau_B \leq t\}}\right] &= L_t^{-1}\left\{\frac{1}{\alpha}\mathbb{E}^Q\left[e^{-\alpha\tau_B}\right]\right\}, \\ \mathbb{E}^Q\left[e^{-\zeta\tau_B}\mathbf{1}_{\{\tau_B \leq t\}}\right] &= L_t^{-1}\left\{\frac{1}{\zeta+\alpha}\mathbb{E}^Q\left[e^{-(\zeta+\alpha)\tau_B}\right]\right\}, \end{aligned}$$

which is related to the computation of the conversion probability and the conversion value in Section 5.3. Note that the relevant expectations under the stock price measure Q^* can be obtained similarly by replacing θ by $\theta + \rho\eta\sigma/\kappa$. As a result, the evaluation of

the CoCo bond can be performed by inverting the Laplace transform numerically. For practical implementation, we propose the use of the Fortet method which is easy-to-program and efficient in computation. The Fortet method can be extended to compute the joint density of the stock price and the first-passage-time (x_τ, τ), which is useful when we need to tackle floored payoff or other additional features.

Fortet Method

The Fortet method can be summarized as follows: we discretize the time into n equal intervals with $t_j = jT/n = j\Delta t$ for $j = 1, 2, ..., n$. The relevant quantities for the pricing of CoCo can be obtained as:

$$Q\left(\tau_B \leq t_j\right) = \sum_{i=1}^{j} q_i, \qquad \mathbb{E}^Q\left[e^{-\zeta\tau_B}\mathbf{1}_{\{\tau_B \leq t_j\}}\right] = \sum_{i=1}^{j} e^{-\zeta t_i} q_i, \tag{5.8}$$

where

$$q_i = \mathbb{P}\left(\tau \in (t_{i-1}, ti]\right), \quad i = 1, 2, ..., n,$$

are obtained by the following recursion

$$\begin{aligned} q_1 &= N\left[a\left(t_1\right)\right], \qquad q_i = N\left[a\left(t_i\right)\right] - \sum_{j=1}^{i-1} q_j N\left[b\left(t_i, t_j\right)\right], \\ a\left(t\right) &= \frac{y_B - M\left(t,0\right)}{S\left(t,0\right)}, \qquad b\left(t,s\right) = \left.\frac{y_B - M\left(t,s\right)}{S\left(t,s\right)}\right|_{y_s = y_B}, \end{aligned}$$

with

$$M\left(t,s\right) = y_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right), \qquad S^2\left(t,s\right) = \frac{\eta^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right).$$

The same procedure applies to the calculation of the relevant quantities under the stock price measure Q^*. Since the idea of the Fortet method is to solve an integral equation by discretizing the time interval, we note that different choices of the approximation scheme that may lead to different forms of the recursive formula. Here, we follow Coculescu et al. (2008) and present the right-point scheme although one can adopt alternative discretization schemes in order to achieve higher convergence (see Appendix).

5.4.4 State-dependent intensity

Under the general hybrid modeling framework, it is possible to incorporate a state-dependent intensity. It is expected that the likelihood of the JtNV should be inversely

related to the stock price or the capital ratio. One popular choice of an intensity specification, as borrowed from credit risk modelings, is

$$\lambda(S) = a + bS^{-\beta},$$

where $a, b, \beta > 0$ and S is the stock price. More examples can be found in Dyrssen et al. (2014) on the pricing equations in jump-to-default models. When the intensity depends on the stock price, it is less straight-forward to reduce the two-dimensional problem into a one-dimension problem. To this end, we can resort to simulation methods to compute the relevant expectations.

To be specific, we can apply a conditional simulation approach here: (i) simulate the log-stock price path $(x_t)_{t\geq 0}$ and compute the cumulative hazard rate function: (ii) conditional on the stock price process, we can compute the first-passage-time density by using the joint density of (x_τ, τ) as approximated by the two-dimensional Fortet method (see Appendix).

5.5 Numerical Illustration

Given the contractual setup of the CoCo bond (c_i, F and G), we need to calibrate the following model parameters for the bivariate model:

- The stock price volatility (σ) can be estimated from historical time-series data or using implied volatility of equity option.
- Because the capital ratio is only observed on an infrequent basis, we can preset the current capital ratio y_0 and roughly estimate the correlation ρ from historical data.
- The long-term mean level θ and mean-reversion speed κ can be estimated from the historical data of the capital ratio. Gupta et al. (2013) mention that major banks disclose the short-term and long-term targets of the tier-1 capital ratio and suggest one to calibrate a mean-reverting model to this information.
- We can calibrate the pricing model to the market price and estimated the market-implied capital ratio volatility η. In particular, the implied capital ratio volatility can be used to quantify how much conversion risk the traders have priced in.

In this comparative analysis, we adopt the bivariate model with a mean-reverting capital ratio enhanced by a JtNV feature. We set the following model parameters: $r = 0.02$,

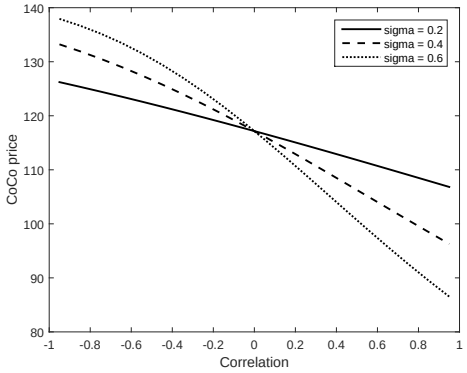
$q = 0.01$, $\sigma = 0.4$, $\eta = 0.3$, $\theta = \ln(0.10)$, $\kappa = 0.2$ and $\rho = 0.75$. We assume that the JtNV intensity is $\lambda = 0.05$ with the jump size is $\gamma = -0.9$, indicating a large write-down in stock price upon a sudden conversion. We consider a CoCo bond with the following specification: 5-year maturity, 10% coupon with the face value of \$100, and the conversion shares is set to be 200 which implies the conversion price is \$0.5. The accounting triggering level is set to be 5% which means $y_B = \ln(0.05)$. The current log stock price is $x_0 = 0.5$ and log capital ratio is taken to be $y_0 = \ln(0.06)$ or $y_0 = \ln(0.10)$.

First-of-all, we recall that the risk-neutral conversion probability under Q depends only on the first-passage-time of the capital ratio, i.e., the marginal dynamics of the capital ratio. The correlation and stock price volatility enter into the pricing formula through the adjustment to the capital ratio dynamics under the stock price measure Q^* when we evaluate the conversion value P_E. This means that the correlation and stock price volatility have no direct impacts on the bond components P_C and P_F.[5] From Lemma 5.2, we see that the conversion value P_E is related to conversion probability under the stock price measure Q^*. When the correlation is positive ($\rho > 0$), the capital ratio dynamics has a higher mean-reverting level due to the adjustment term $\rho\sigma\eta > 0$. This implies a smaller conversion probability under Q^* and hence a smaller conversion value P_E.

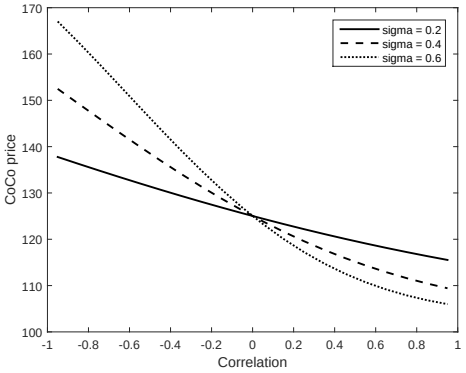
5.5.1 The impact of correlation

Figure 5.2 shows that the CoCo bond price declines as the correlation moves from -1 to 1 under the bivariate framework. When the correlation is positive, the capital ratio moves in tandem with the stock price. This means that the stock price is likely to be cheap when the capital ratio falls and bleaches the accounting threshold, which in turn generates a smaller conversion value P_E. The effect is more prominent when the current capital ratio is close to the threshold as shown in the left panel of Figure 5.2. Conversely, when the correlation is negative, the capital ratio moves in an opposite direction to the stock price. As such, the stock price is likely to be more expensive at an accounting triggered conversion and this implies a higher expected conversion value P_E. The latter phenomenon is more prominent when the current capital ratio is far away from the trigger as shown in the right panel of Figure 5.2.

[5]This is unlike the equity derivative approach in which the conversion time is proxied by the first-passage-time of the stock price to an implied threshold, in which the stock price volatility affects directly the CoCo bond price through the conversion probability.

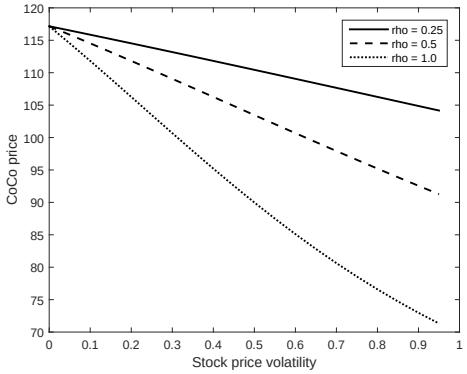


(A) The tier-1 capital ratio = 6%.

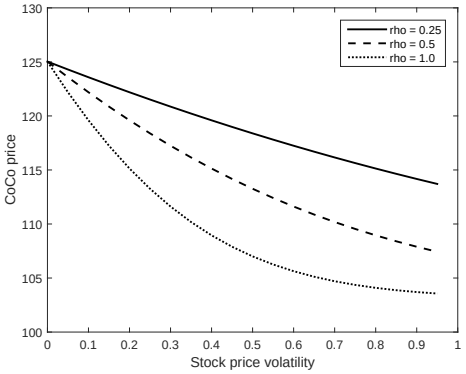


(B) The tier-1 capital ratio = 10%.

FIGURE 5.2: Impact of correlation to the CoCo bond price.



(A) The tier-1 capital ratio = 6%.



(B) The tier-1 capital ratio = 10%.

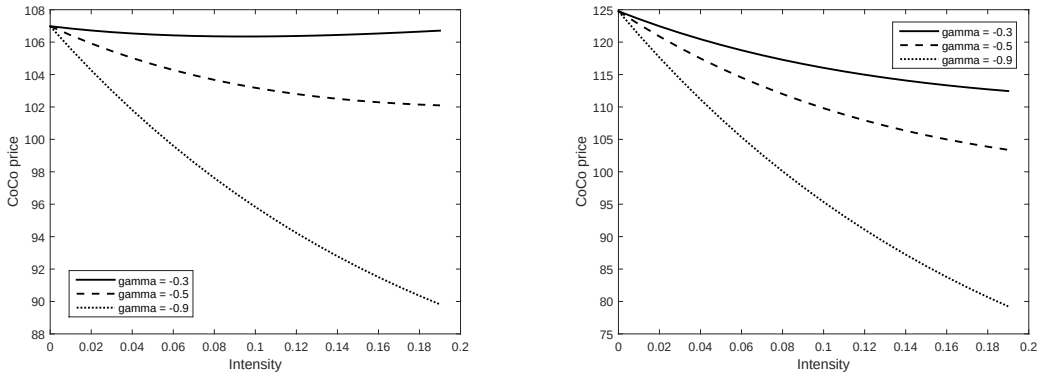
FIGURE 5.3: Impact of stock price volatility to the CoCo bond price.

5.5.2 The impact of stock price volatility

Figure 5.3 shows that the CoCo bond price is decreasing with the stock price volatility at different levels of positive correlation as $\rho = 0.25$, $\rho = 0.5$ and $\rho = 1.0$. When the stock price volatility is high, the CoCo bond is likely to be converted when the stock price declines much below than its current value. As expected, the sensitivity to stock price volatility is higher when the correlation is closer to one. We see that the CoCo bond price is more sensitive to the stock volatility when the capital ratio is near the triggering threshold, suggesting that there is a significant equity component of the CoCo bond.

5.5.3 The impact of intensity

Figure 5.4 reports the sensitivity of the CoCo bond price to the level of JtNV intensity λ at different size of the stock price jump at conversion. We vary the jump size as γ



(A) The tier-1 capital ratio = 6%. (B) The tier-1 capital ratio = 10%.

FIGURE 5.4: Impact of JtNV intensity to the CoCo bond price.

= -0.3, -0.5, -0.9, which means that the write-down due to a JtNV is assumed to be 30%, 50% and 90%, respectively. When the capital ratio is far away from the triggering threshold (right panel), most of the conversion risk of a CoCo bond lies at the possibility of a JtNV with the corresponding write down of the stock price. This can be captured by the added JtNV feature which reduces the value of a CoCo bond as the intensity increases. From a modeling perspective, the added JtNV feature is useful to build in the risk premium of a CoCo bond when the conversion probability generated by the capital ratio diffusion is too small.

It is intriguing to look closer to the case of $\gamma = -0.3$ in the left panel of Figure 5.4. When the capital ratio is close to the triggering threshold, a sudden jump to the non-viability state may actually improves the value of a CoCo bond if the stock price write-down is not large. This is because a sudden JtNV conversion well before the stock price declining to a low value might actually provide a higher recovery which is beneficial to the CoCo bond investor. The non-monotonic sensitivity to the JtNV intensity suggests that one should be careful in setting the jump size when calibrate to the market prices of CoCo bonds.

5.6 Summary

In this chapter, we study the pricing of contingent convertibles based on an equity-credit hybrid approach. We propose a bivariate modeling of the stock price and the tier-1 capital ratio which is suitable for the pricing of contingent convertible bonds with an accounting trigger. Furthermore, we introduce an enhanced jump-to-non-viability feature to incorporate a more realistic scenario in which a CoCo bond is triggered by a rare event such as a large capital write-down of the issuance bank. In the model implementation, we derive the closed-form pricing formula when the log capital-ratio follows

a driftless Brownian motion and we demonstrate the application of the Fortet method to compute the first-passage-time distribution for the mean-reverting process. The numerical example shows that the bivariate model enhanced by a jump-to-non-viability feature provide a rich and flexible framework for the pricing and risk-management of CoCo bonds.

5.7 Appendix

5.7.1 First-passage-time problem for the bivariate process

For the ease of exposition, we use the notation $\mathbb{P}(\xi \in dx) \triangleq \Pr(\xi \in [x, x+dx))$ for a random variable ξ. Take $X_t = (x_t, y_t)$ to be a bivariate Gaussian process taking values in $\mathbb{R}^2$, with x_t is the observable marker process and y_t is the latent process. We are interested in the first-passage-time (FPT) of the latent process to a lower threshold $y_B < y_0$ as

$$\tau = \inf(t \geq 0; y_t = y_B).$$

On the event $\{\tau \in dt\}$, by continuity we see that $y_t = y_B$. Hence, we only need to compute the joint density of (x_τ, τ) instead of (y_τ, x_τ, τ). The joint density of (x_τ, τ) can be formulated as

$$\mathbb{P}(x_\tau \in dx, \tau \in dt) = q(x,t)\,dxdt.$$

In terms of distribution, we have

$$\mathbb{P}(x_\tau \leq x, \tau \leq t) = \int_0^t \int_{-\infty}^{x} q(x', t')\,dx'dt'.$$

By sending $x \to \infty$, we obtain the FPT distribution

$$\mathbb{P}(\tau \leq t) = \int_0^t \int_{-\infty}^{\infty} q(x', t')\,dx'dt'.$$

Moreover, the FPT density can be formulated as

$$\mathbb{P}(\tau \in dt) = g(t)dt.$$

By conditional probability, we see that

$$\mathbb{P}(x_\tau \in dx, \tau \in dt) = \mathbb{P}(x_\tau \in dx | \tau \in dt)\,\mathbb{P}(\tau \in dt),$$

and

$$\mathbb{P}(x_\tau \in dx | \tau \in dt) = \frac{\mathbb{P}(x_\tau \in dx, \tau \in dt)}{\mathbb{P}(\tau \in dt)},$$

which is the conditional distribution of x_t when the hitting time is at $t = \tau$. A similar bivariate first-passage-time problem with one-side threshold has been tackled in Longstaff and Schwartz (1995), Collin-Dufresne and Goldstein (2001) and Coculescu et al. (2008) for structural credit risk modeling, and Bernard et al. (2008) for exotic barrier option pricing. Next, let us review the closed-form solution for the case of Brownian model and the numerical method for the case of mean-reverting model.

Brownian model

This exposition follows Kijima et al. (2009a). For $t \geq 0$, the bivariate process can be equivalently written as

$$\begin{aligned} x_t &= x_0 + \beta t + \rho\sigma W_t^1 + \sigma\sqrt{1-\rho^2}W_t^2, \\ y_t &= y_0 + \alpha t + \eta W_t^1, \end{aligned}$$

in which W_t^1 and W_t^2 are independent Brownian motions. We can see that

$$\begin{aligned} \mathbb{P}\left(x_t \in dx, \tau \in dt\right) &= \mathbb{P}\left(x_t \in dx \middle| \tau \in dt\right)\mathbb{P}\left(\tau \in dt\right) \\ &= \mathbb{P}\left(x_0 + \beta t + \rho\sigma W_t^1 + \sigma\sqrt{1-\rho^2}W_t^2 \in dx \middle| \tau \in dt\right)\mathbb{P}\left(\tau \in dt\right). \end{aligned}$$

On the event $\{\tau \in dt\}$, we have $y_t = y_B$ such that $W_t^1 = \left(y_B - y_0 - \alpha t\right)/\eta$, and hence

$$\begin{aligned} & \mathbb{P}\left(x_0 + \beta t + \rho\sigma W_t^1 + \sigma\sqrt{1-\rho^2}W_t^2 \in dx \middle| \tau \in dt\right) \\ = & \mathbb{P}\left(x_0 + \beta t + \frac{\rho\sigma}{\eta}\left(y_B - y_0 - \alpha t\right) + \sigma\sqrt{1-\rho^2}W_t^2 \in dx \middle| \tau \in dt\right) \\ = & \mathbb{P}\left(x_0 + \beta t + \frac{\rho\sigma}{\eta}\left(y_B - y_0 - \alpha t\right) + \sigma\sqrt{1-\rho^2}W_t^2 \in dx\right), \end{aligned}$$

in which the second equality follows from the fact that W_t^2 and τ are independent. In this case, it is straight-forward to reduce the two-dimensional FPT problem into a one-dimensional problem. Indeed, we can express the joint density of (x_τ, τ) as

$$q\left(x,t\right) = \pi\left(x,t\right)g\left(t\right),$$

where

$$\pi\left(x,t\right) = \frac{1}{\sqrt{2\pi\left(1-\rho^2\right)\sigma^2 t}}\exp\left\{-\frac{1}{2\left(1-\rho^2\right)\sigma^2 t}\left(x - x_0 - \beta t - \rho\frac{\sigma}{\eta}\left(y_B - y_0 - \alpha t\right)\right)\right\},$$

and

$$g(t) = \frac{y_0 - y_B}{\sqrt{2\pi\eta^2 t^3}}\exp\left[-\frac{\left(y_B - y_0 - \alpha t\right)^2}{2\eta^2 t}\right],$$

are the Gaussian density and the inverse Gaussian density respectively. Then, it is easy to show that

$$\begin{aligned}\mathbb{P}(\tau \leq t) &= \int_0^t \int_{-\infty}^{\infty} q(x', t')\, dx' dt' \\ &= N\left[\frac{y_B - y_0 - \alpha t}{\eta\sqrt{t}}\right] + e^{\frac{2\beta}{\eta^2}(y_B - y_0)} N\left[\frac{y_B - y_0 + \alpha t}{\eta\sqrt{t}}\right],\end{aligned}$$

which is just the standard FPT distribution as in Black and Cox (1976).

Mean-reverting model

The joint dynamics of (x_t, y_t) conditional on the information up to $s < t$ is

$$\begin{aligned}x_t &= x_s + \alpha(t-s) + \int_s^t \rho\sigma dW_u^1 + \int_s^t \sigma\sqrt{1-\rho^2} dW_u^2, \\ y_t &= y_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right) + \int_s^t \eta e^{-\kappa(t-u)} dW_u^1.\end{aligned}$$

In the following, we demonstrate the integral equation approach, known as the Fortet method, to compute the relevant quantities for a mean-reverting process.

One-dimensional Fortet method.

Firstly, let us focus on the marginal dynamics of the capital ratio y_t. Denote the transitional probability density for the process y_t conditional on $s < t$ as

$$\mathbb{P}\left\{ y_t \in dy \middle| y_s \in dy' \right\} = f(y, t; y', s)\, dy.$$

By the continuity of the process, we have

$$f(y, t; y_0, 0) = \int_0^t g(s)\, f(y, t; y_B, s)\, ds.$$

Integrate y on both side over $(-\infty, y_B]$, we arrive the Fortet's equation as

$$N\left[\frac{y_B - \mu(t,0)}{\Sigma(t,0)}\right] = \int_0^t g(s)\, N\left[\frac{y_B - \mu(t,s)}{\Sigma(t,s)}\right]\Bigg|_{y_s = y_B} ds, \tag{5.9}$$

where

$$\mu(t,s) = y_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right), \quad \Sigma(t,s) = \frac{\eta^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right).$$

By taking

$$a(t) = \frac{y_B - \mu(t,0)}{\Sigma(t,0)}, \qquad b(t,s) = \left.\frac{y_B - \mu(t,s)}{\Sigma(t,s)}\right|_{y_s = y_B},$$

we can express (5.9) as

$$N\left[a\left(t\right)\right]=\int_{0}^{t}g\left(s\right)N\left[b\left(t,s\right)\right]ds,$$

which can be seen as a Volterra integral equation of the first kind. To solve the equation numerically, let us apply the right-point scheme and evaluate the integral on the discrete time intervals $t\rightarrow t_{j}=j\Delta t$, $j=1,...,m$, with Δt is the uniform grid size. Hence, we have

$$N\left[a\left(t_{j}\right)\right]=\sum_{h=1}^{j}q_{h}N\left[b\left(t_{j},t_{h}\right)\right]$$

with the approximation

$$q_{j}=\mathbb{P}\left(\tau\in\left(t_{j-1},t_{j}\right]\right)\approx g\left(t_{j}\right)\Delta t,\qquad j=1,2,...,m.$$

Then, we can derive the recursive relationship

$$\begin{aligned}q_{1} &= N\left[a\left(t_{1}\right)\right],\\ q_{j} &= N\left[a\left(t_{2}\right)\right]-\sum_{h=1}^{j-1}q_{h}N\left[b\left(t_{j},t_{h}\right)\right],\qquad j=2,3,...,m.\end{aligned}$$

This is the formula as presented in Longstaff and Schwartz (1995) and Coculescu et al. (2008). Coculescu et al. (2008) mention that adopting the right-point scheme means that default (conversion) may only occur at the ends of the subperiods. Hence, the right-point scheme delivers a clearer probabilistic interpretation of the term q_i which makes the extension to higher dimensional problem more straight-forward.

To achieve higher convergence, one can adopt a composite trapezium rule to solve the integral equation. By similar arguments, it can be shown that

$$\begin{aligned}q_{1} &= 2N\left[a\left(t_{1}\right)\right],\\ q_{j} &= 2\left\{N\left[a\left(t_{2}\right)\right]-\sum_{h=1}^{j-1}q_{h}N\left[b\left(t_{j},t_{h}\right)\right]\right\},\qquad j=2,3,...,m,\end{aligned}$$

with the approximation $q_{j}\approx g\left(t_{j}\right)\Delta t$ for $j=0,1,2,...,m$ and noting that $q_{0}=0$. In this case, the probability interpretation of q_i is less clear although it provides a better estimate of $g(t_i)$. Linz (1969) discusses several numerical discretization for solving the integral equation and the accuracy of the rectangular scheme, trapezoidal scheme and mid-point scheme.

Two-dimensional Fortet method.

The one-dimensional Fortet method is very efficient when the first-passage-time of y_t is concerned. In some applications, we are interested to compute the joint density of (x_τ, τ) as

$$\mathbb{P}\{x_\tau \in dx, \tau \in dt\} = q(x,t)\,dxdt.$$

Denote the transitional probability density for the bivariate process (y_t, x_t) conditional on $s < t$ as

$$\mathbb{P}\left\{ y_t \in dy, x_t \in dx \middle| y_s \in dy', x_s \in dx' \right\} = f\left(y, x, t; y', x', s\right) dydx.$$

By the continuity of the process (y_t, x_t), we have

$$f(y, x, t; y_0, x_0, 0) = \int_0^t \int_{-\infty}^{\infty} g\left(x', s\right) f\left(y, x, t; y_B, x', s\right) dx'ds.$$

Integrating y on both sides over $(-\infty, y_B]$, we arrive the extended Fortet's equation as

$$\Phi(x,t) = \int_0^t \int_{-\infty}^{\infty} g\left(x', s\right) \psi\left(x, t; x', s\right) dx'ds, \tag{5.10}$$

where

$$\begin{aligned} \Phi(x,t) &= \int_{-\infty}^{y_B} f(y, x, t; y_0, x_0, 0)\, dy, \\ \psi\left(x, t; x', s\right) &= \int_{-\infty}^{y_B} f\left(y, x, t; y_B, x', s\right) dy, \end{aligned}$$

are available in closed-form for a bivariate Gaussian process. Since $y_0 > y_B$, it can be checked that

$$\lim_{t\to 0} \Phi(x,t) = 0, \qquad \lim_{t\to s} \psi\left(x, t; x', s\right) = \delta\left(x - x'\right), \tag{5.11}$$

where $\delta(\cdot)$ is the Dirac-delta function. In order to solve the two-dimensional integral equation (5.10), we discretize the domain for x and t using rectangular grids with the right-point scheme as:

$$\begin{aligned} x_i &= x_{lb} + i\Delta x, \qquad i = 1, 2, ..., n, \\ t_j &= j\Delta t, \qquad\quad\ \ j = 1, 2, ..., m, \end{aligned}$$

so that we can express

$$\Phi(x_i, t_j) = \sum_{h=1}^{j} \sum_{k=1}^{n} q(x_k, t_h)\, \psi(x_i, t_j; x_k, t_h),$$

by taking the approximation

$$\begin{aligned} q(x_i, t_j) &= \mathbb{P}(x \in (x_{i-1}, x_i], \tau \in (t_{j-1}, t_j]) \\ &\approx g(x_i, t_j) \Delta x \Delta t. \end{aligned}$$

For $j = 1$, we see that

$$\Phi(x_i, t_1) = \sum_{k=1}^{n} q(x_k, t_1) \psi(x_i, t_1; x_k, t_1) = (\Delta x)^{-1} q(x_i, t_1),$$

by applying the property in (5.11). Hence, we have

$$q(x_k, t_1) = \Delta x \Phi(x_i, t_1).$$

For $j > 1$, we can re-write

$$\Phi(x_i, t_j) = \sum_{k=1}^{n} q(x_k, t_j) \psi(x_i, t_j; x_k, t_j) + \sum_{h=1}^{j-1} \sum_{k=1}^{n} q(x_k, t_h) \psi(x_i, t_j; x_k, t_h).$$

By (5.11), the first term on the RHS gives $(\Delta x)^{-1} q(x_i, t_j)$. As a result, we have

$$q(x_i, t_j) = \Delta x \left[\Phi(x_i, t_j) - \sum_{h=1}^{j-1} \sum_{k=1}^{n} q(x_k, t_h) \psi(x_i, t_j; x_k, t_h) \right],$$

which results in a recursive relationship for $j = 2, 3, ..., m$. Note that the composite trapezium rule can be incorporated by a similar modification as in the case of one-dimensional Fortet method.

We also need the closed-form expression for Φ and ψ. To this end, we see that $y_t|_{x_t, \mathcal{F}_s} \sim N\left(\mu(t; s), \Sigma^2(t; s)\right)$ is Gaussian in which the conditional moments can be obtained by the projection theorem as

$$\begin{aligned} \mu(t; s) &\triangleq \mathbb{E}_s[y_t | x_t] = \mathbb{E}_s[y_t] + \frac{Cov_s(y_t, x_t)}{Var_s[x_t]} [x_t - \mathbb{E}_s[x_t]], \\ \Sigma^2(t; s) &\triangleq Var_s[y_t | x_t] = Var_s[y_t] - \frac{Cov_s(y_t, x_t)^2}{Var_s[x_t]}, \end{aligned}$$

in which the moments are readily obtained as

$$\begin{aligned} &\mathbb{E}_s[x_t] = x_s + \alpha(t - s), \quad \mathbb{E}_s[y_t] = y_s e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right), \\ &Var_s[x_t] = \sigma^2(t - s), \quad Var_s[y_t] = \frac{\eta^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right), \\ &Cov_s(y_t, x_t) = \rho\sigma\eta\left(\frac{1 - e^{-\kappa(t-s)}}{\kappa}\right). \end{aligned}$$

By conditional probability, we have

$$f\left(y_t, x_t, t; y_s, x_s, s\right) = f\left(x_t, t; x_s, s\right) f\left(y_t, t; y_s, x_s, s \middle| x_t\right),$$

where

$$\begin{aligned} f\left(x_t, t; x_s, s\right) &= \frac{1}{\sqrt{2\pi\sigma^2\left(t-s\right)}} \exp\left(-\frac{\left(x_t - x_s - \alpha\left(t-s\right)\right)^2}{2\sigma^2\left(t-s\right)}\right), \\ f\left(y_t, t; y_s, x_s, s \middle| x_t\right) &= \frac{1}{\sqrt{2\pi\Sigma^2\left(t;s\right)}} \exp\left(-\frac{\left(y_t - \mu\left(t;s\right)\right)^2}{2\sigma^2\Sigma^2\left(t;s\right)}\right). \end{aligned}$$

By direct integration, we obtain the closed-form expression for $\Phi\left(x,t\right)$ and $\psi\left(x,t;x',s\right)$ as

$$\begin{aligned} \Phi\left(x,t\right) &= f\left(x_t, t; x_0, 0\right) N\left[\frac{y_B - \mu\left(t;0\right)}{\Sigma\left(t;0\right)}\right], \\ \psi\left(x,t;x',s\right) &= f\left(x_t, t; x_s, s\right) N\left[\frac{y_B - \mu\left(t;s\right)}{\Sigma\left(t;s\right)}\right]\Bigg|_{y_s = y_B}. \end{aligned}$$

This completes the algorithm. Finally, one can check the numerical consistency of the extended Fortet method and the one-dimensional Fortet method by integrating out x (i.e., summing up the index i) and compute the FPT density or distribution.

5.7.2 Proof of Lemma 5.1

Decompose the event $\{\tau_B > \tau_R\}$ and $\{\tau_R > \tau_B\}$, we have

$$Q\left(\tau \leq T\right) = \mathbb{E}^Q\left[\mathbf{1}_{\{\tau_B \wedge \tau_R \leq T\}}\right] = \mathbb{E}^Q\left[\mathbf{1}_{\{\tau_R \leq T\}}\mathbf{1}_{\{\tau_B > \tau_R\}}\right] + \mathbb{E}^Q\left[\mathbf{1}_{\{\tau_B \leq T\}}\mathbf{1}_{\{\tau_R > \tau_B\}}\right].$$

The first term represents the conversion probability when the regulatory trigger occurs before the accounting trigger, which can be computed by iterated expectation as

$$\begin{aligned} \mathbb{E}^Q\left[\mathbf{1}_{\{\tau_R \leq T\}}\mathbf{1}_{\{\tau_B > \tau_R\}}\right] &= \mathbb{E}^Q\left[\mathbb{E}^Q\left[\mathbf{1}_{\{\tau_R \leq T\}}\mathbf{1}_{\{\tau_B > \tau_R\}}\middle| \tau_R = u\right]\right] \\ &= \int_0^T \mathbb{E}^Q\left[\lambda\left(X_u\right) e^{-\int_0^u \lambda(X_s)ds}\mathbf{1}_{\{\tau_B > u\}}\right] du. \end{aligned}$$

The second term represents the conversion probability when the accounting trigger occurs before the regulatory trigger, which is obtained as

$$\begin{aligned} \mathbb{E}^Q\left[\mathbf{1}_{\{\tau_R \leq T\}}\mathbf{1}_{\{\tau_R > \tau_B\}}\right] &= \mathbb{E}^Q\left[\mathbb{E}^Q\left[\mathbf{1}_{\{\tau_B \leq T\}}\mathbf{1}_{\{\tau_R > \tau_B\}}\middle| \tau_B = u\right]\right] \\ &= \mathbb{E}^Q\left[e^{-\int_0^{\tau_B}\lambda(X_u)du}\mathbf{1}_{\{\tau_B \leq T\}}\right]. \end{aligned}$$

5.7.3 Proof of Lemma 5.2

The proof follows from Lemma 11.6.1 and Theorem 11.6.2 in Shreve (2008). From the adjusted stock price process $\widetilde{S}_t$, we can decompose the Radon-Nikodym density as

$$Z_t = Z_t^c Z_t^J,$$

where

$$Z_t^c = \exp\left(\int_0^t \sigma dW_s^1 - \int_0^t \frac{1}{2}\sigma^2 ds\right), \qquad Z_t^J = \exp\left(-\gamma\Lambda_t\right)(1+\gamma)^{N_t},$$

correspond to the change-of-measure for the continuous path and jump path respectively.

As the Brownian motion is only affected by the change-of-measure Z_t^c, by the Girsanov theorem we have

$$dB_t^1 = dW_t^1 - \sigma dt,$$

and hence

$$dy_t = \alpha\left(y_t\right)dt + \eta\left(\sqrt{1-\rho^2}dW_t^{\perp} + \rho dW_t^1\right) = \left[\alpha\left(y_t\right) + \rho\sigma\eta\right]dt + \eta dB_t^2,$$

by taking $dB_t^2 = \sqrt{1-\rho^2}dW_t^{\perp} + \rho dB_t^1$ with $\left\langle dB_t^1, dB_t^2\right\rangle = \rho dt$.

On the other hand, the Poisson process is affected only by the change-of-measure Z_t^J as

$$Z_t^J = \exp\left(-\gamma\Lambda_t\right)(1+\gamma)^{N_t} = e^{(\lambda-\lambda^*)t}\left(\frac{\lambda^*}{\lambda}\right)^{N_t},$$

in which the second equality comes from the change-of-measure for a Poisson process with constant intensity (Shreve, 2008). Hence, we can take

$$\lambda^* = (1+\gamma)\lambda$$

as the intensity under the stock price measure Q^*.

5.7.4 Proof of Proposition 5.4

We can perform the integration-by-part as

$$\begin{aligned}\int_0^T \lambda^* e^{-(\lambda^*+q)u} Q^*(\tau_B > u)\, du &= -\frac{\lambda^*}{\lambda^*+q}\int_0^T Q^*(\tau_B > u)\, d\left(e^{-(\lambda^*+q)u}\right) \\ &= \frac{\lambda^*}{\lambda^*+q}\left[1 - e^{-(\lambda^*+q)T} Q^*(\tau_B > T)\right] \\ &\quad + \frac{\lambda^*}{\lambda^*+q}\int_0^T e^{-(\lambda^*+q)u}\left[-\frac{\partial Q^*}{\partial u}(\tau_B > u)\right] du,\end{aligned}$$

since $g^{Q^*}(t; y_B) = -\partial_t Q^*(\tau_B > t)$ is the density of τ_B, we have

$$\int_0^T e^{-(\lambda^*+q)u}\left[-\frac{\partial Q^*}{\partial u}(\tau_B > u)\right] du = \mathbb{E}^{Q^*}\left[e^{-(\lambda^*+q)\tau_B}\mathbf{1}_{\{\tau_B \le T\}}\right],$$

Hence, we have

$$\mathbb{E}^{Q^*}\left[e^{-q\tau}\mathbf{1}_{\{\tau\le T\}}\right] = \frac{\lambda^*}{\lambda^*+q}\left[1 - e^{-(\lambda^*+q)T} Q^*(\tau_B > T)\right] + \frac{q}{\lambda^*+q}\mathbb{E}^{Q^*}\left[e^{-(\lambda^*+q)\tau_B}\mathbf{1}_{\{\tau_B\le T\}}\right].$$

5.7.5 Proof of Proposition 5.5

For $\zeta \ge 0$, we have

$$\mathbb{E}^{Q^*}\left[e^{-\zeta\tau}\mathbf{1}_{\{\tau\le T\}}\right] = \int_0^T e^{-\zeta t} g^{Q^*}(t; y_B)\, dt,$$

where

$$g^{Q^*}(t; y_B) = \frac{y_0 - y_B}{\sqrt{2\pi\eta^2 t^3}} \exp\left[-\frac{(y_B - y_0 - \tilde{\alpha}t)^2}{2\eta^2 t}\right],$$

with $\tilde{\alpha} = \alpha + \rho\sigma\eta$ is the first-passage-time density under the stock price measure Q^*. By completing square we see that

$$\mathbb{E}^{Q^*}\left[e^{-\zeta\tau}\mathbf{1}_{\{\tau\le T\}}\right] = e^{\frac{\tilde{\alpha}-\Gamma}{\eta^2}(y_B - y_0)}\int_0^T \frac{y_0 - y_B}{\sqrt{2\pi\eta^2 t^3}} \exp\left[-\frac{(y_B - y_0 - \Gamma t)^2}{2\eta^2 t}\right] dt,$$

where $\Gamma = \sqrt{\tilde{\alpha}^2 + 2\zeta\eta^2}$. By direct integration, we have

$$I = e^{\delta^-(y_B - y_0)} N\left[\frac{y_B - y_0 - \Gamma t}{\eta\sqrt{t}}\right] + e^{\delta^+(y_B - y_0)} N\left[\frac{y_B - y_0 + \Gamma t}{\eta\sqrt{t}}\right],$$

where $\delta^{\pm} = \dfrac{\tilde{\alpha} \pm \Gamma}{\eta^2}$.

Chapter 6

Asymptotic Expansion for Multifactor Heston Model

6.1 Introduction

For the last two decades, the stochastic volatility model of Heston (1993) has been one of the most popular choices in the modeling of price dynamics of various asset classes, including equity price, foreign exchange rate, and interest rate. In the Heston model, one assumes that the asset price follows a stochastic volatility model in which the variance term of the stock price is driven by a Cox-Ingersoll-Ross (CIR) process and can be correlated with the asset price itself. The Heston model provides a succinct description to a number of important empirical features in the dynamics of asset price, such as the leverage effect, the mean-reversion nature and the clustering of volatility. In terms of derivatives pricing, the model has gained popularity on trading desks, given its ability to manage the implied volatility smile and its good analytical tractability in the pricing of standard European-type options. Nevertheless, the one-factor dynamics of volatility in Heston (1993) has a number of drawbacks. Firstly, it does not provide enough flexibility to model both the short-term and long-term implied volatility smile simultaneously. Thus, the level and slope of the smile generated by the model cannot be independently determined. This fact often forces analysts to adapt two different sets of model parameters to separately price and risk-manage the short-term and long-term options. Such a remedy could misprice exotic derivatives that are sensitive to the realized path of volatility. On the other hand, the assumption of constant model parameters implies certain restrictions on the shape of the model-implied volatility surface. For example, when the correlation between the asset price and stochastic volatility is constant, the term structure of the volatility skew is governed by the speed of mean-reversion of

the volatility process. This restricts the implied volatility surface to flattenning out in the long-run, making it difficult to capture the market scenario in which the long-term skewness is persistent. Some hedge funds have reportedly taken advantage of such mispricing in long-term options and generated significant profits during the recent financial crisis. These shortcomings of the Heston model have raised the question of whether a more realistic description of the volatility dynamics could be obtained by increasing the number of factors of the variance process, or by relaxing the assumption of constant model parameters.

In view of these concerns, there are a growing number of papers that consider the multifactor extension of the Heston model in derivatives pricing. In one such paper, Fonseca et al. (2008) propose a multifactor Heston model based on the Wishart process, wherein they consider the option pricing in the foreign exchange market. Meanwhile, Christoffersen et al. (2009) show that a two-factor Heston model performs much better than a one-factor model in capturing the dynamics of the implied volatility of the S&P 500 index options. This is because the two-factor model allows for flexible modeling of the volatility surface, such that the level and slope of the volatility smile can be independently determined with the additional degree of freedom. They also find that the estimated variance factors can be identified as a strongly mean-reverting short-term variance factor and a slowly mean-reverting long-term factor. In addition, Li and Zhang (2010) use a non-parametric approach to analyze the index option dataset and verify the conclusion in Christoffersen et al. (2009). They find that one needs to use at least two factors in order to sufficiently capture the dynamics of implied volatility in both the time-series and cross-sectional dimensions. In fact, in proposing a class of no-arbitrage variance curve models with multiple stochastic volatility factors, Buehler (2006) finds that one needs a two-factor model to sufficiently capture the term structure of variance swaps written on major stock market indices. These observations are also consistent with the series of works by Fouque et al. (2000, 2003, 2004), which propose the multi-scale nature of stochastic volatility. Fouque and Lorig (2011) propose an extension to the one-factor Heston model by adding an additional fast mean-reverting component. Their results also show that such an extension allows a significant improvement in the fitting of volatility smile for index options. Meanwhile, Bergomi (2008) highlights the fact that the one-factor Heston model could lead to the mispricing of exotic derivatives, such as forward-starting options, cliquets and variance swaps. He suggests that a properly designed volatility smile model for the pricing and risk management of exotic derivatives should have separate controls of (i) the term structure of volatility, (ii) short-term volatility skew, and (iii) the correlation between the spot price and volatility.

As a mathematical analysis tool to study the behavior of volatilities under various dynamical assumptions, a number of papers have adopted asymptotic approaches based

on perturbations and Malliavin calculus. For short-dated options, for example, Hagan et al. (2003) apply singular perturbations to obtain the asymptotic formula of the SABR model, while Henry-Labordere (2005) employs a heat-kernel expansion to the mean-reverting SABR model. On the other hand, Alós et al. (2007) use Malliavin calculus to study the short-term behavior of implied volatility for jump-diffusion models with stochastic volatility. These expansions prove to be accurate and useful for short-term (near expiry) options. The asymptotic expansion approach can also be applied to study implied volatility of longer maturity. Using a partial differential equation (PDE) approach, Fouque et al. (2000) propose a singular perturbation with respect to a fast mean-reverting parameter of the stochastic volatility and, in a later study, Fouque et al. (2004) extend the approach to the case of time-periodic coefficients that capture the maturity cycles. Fukasawa (2011) also derives several well-known results of regular and singular perturbations on stochastic volatility models using the martingale expansion techniques. Meanwhile, Antonelli and Scarlatti (2009) consider an expansion with respect to the correlation coefficient and develop a power series formula for option prices. Now, for small volatility expansion, Osajima (2007) applies Malliavin calculus and the large deviation theory to study a dynamic SABR model in which the parameters are time-dependent, while Takahashi and Yamada (2012) develop the asymptotic expansion around the Black-Scholes model for a general class of multi-dimensional stochastic volatility model with jump diffusion. Asymptotic expansion has also been found to be very efficient in other areas of derivatives pricing. For example, Tanaka et al. (2010) use the Gram-Charlier expansion to derive asymptotic approximation for interest rate derivatives, Papageorgiou and Sircar (2009) use singular perturbations to price single-name and multi-name credit derivatives under a stochastic volatility environment, and more recently, Bayraktar and Yang (2011) use similar techniques for equity-credit hybrid modeling. Benhamou et al. (2010) employ Malliavin calculus to develop a fast and accurate approximation formula of option prices under the one-factor Heston model with time-dependent parameters. By the asymptotic expansion with respect to the volatility of volatility, they show that the European put option price can be approximated by the Black-Scholes formula, with a number of correction terms related to the Greeks of the option.

In this chapter, we extend the results of Benhamou et al. (2010) in a simple way and, thus, develop the approximation formula under the general multifactor Heston model with time-dependent parameters. Our aim is to model the implied volatility surface more realistically, but, with reasonable computational intensities. We demonstrate the ability of the multifactor model to reproduce various shapes of the implied volatility surface through the numerical illustration and the calibration. The contributions are twofold: the theoretical expansion of the literature and the numerical analysis. In the

theoretical aspect, we find that the expansion terms can be expressed as a sum of the expansion terms obtained in Benhamou et al. (2010) plus a new term that captures the interaction between different variance factors. The approximation formula allows one to study the effect of multifactor extension and time-dependent parameters in a simple and unified framework. The formula under constant parameters can be explicitly computed and the incorporation of time-dependent parameters is straightforward. Thus, it can also be shown that the approximation formula is fast and efficient and significantly reduces the computational time for calibration to market data. A fast and efficient approximation formula is useful when one has to compute a large number of option prices, such as econometrics estimation using option data and evaluation of large-scale portfolio risk. For the numerical analysis, we study the accuracy of the formula under different parameter settings and calibrate the two-factor model to the real data of index options and variance swaps for the S&P 500 and the Nikkei 225 index. We find that it is possible to distinguish a short-term and long-term variance factor from the implied volatility surface and variance swap rates on these indices. Moreover, the two-factor model is able to reproduce the shapes of the implied volatility surface during various market scenarios.

6.2 Multifactor Heston Model

6.2.1 Mathematical formulation

We fix a maturity date T and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$, where $\mathbb{P}$ denotes the risk-neutral measure with a deterministic interest rate r_t. We assume that on the filtered probability space, n-pairs of correlated Brownian motions

$$\left\{\left(W_t^1, B_t^1\right), \left(W_t^2, B_t^2\right), ..., \left(W_t^n, B_t^n\right)\right\},$$

are equipped with a correlation structure

$$d\langle W^i, W^j\rangle_t = d\langle B^i, B^j\rangle_t = \delta_{ij}dt, \quad d\langle W^i, B^j\rangle_t = \delta_{ij}\rho_{it}dt, \quad i,j \in \{1,2,...,n\},$$

where δ_{ij} is the Kronecker delta, and the σ-algebra $\mathcal{F}_t$ is generated by these Brownian motions up to time t. X_t is the log-forward price with a fixed maturity T of a security. X_t and the variance factors v_{it}, $i = 1, 2, ..., n$, are supposed to satisfy the following

system of stochastic differential equations (SDEs)

$$X_t = x_0 + \sum_{i=1}^{n} \left[\int_0^t \sqrt{\upsilon_{is}} dW_s^i - \frac{1}{2} \int_0^t \upsilon_{is} ds \right],$$
$$\upsilon_{it} = \upsilon_{i0} + \int_0^t \kappa_i(\theta_{is} - \upsilon_{is}) ds + \int_0^t \xi_{is} \sqrt{\upsilon_{is}} dB_s^i, \tag{6.1}$$

where κ_i is the mean-reversion speed, θ_{is} is the mean-reversion level and ξ_{is} is the volatility-of-volatility (Vol-of-Vol) of the i-th variance factor. For each variance factor υ_{it}, we assume that the parameters $(\kappa_i, \upsilon_{i0})$ are positive constants, while we allow the parameters $(\theta_{it}, \xi_{it}, \rho_{it})$ to be time-dependent (deterministic). We write $\upsilon_t = \sum_{i=1}^n \upsilon_{it}$. In order to guarantee the positivity of the variance factors and the nondegeneracy of X_t, we set out the following assumptions on the model parameters throughout the derivation in Sections 6.2 and 6.3.

Assumption I

$$\inf_{t\in[0,T]} \xi_{it} > 0, \quad \sup_{t\in[0,T]} |\rho_{it}| < 1, \quad \inf_{t\in[0,T]} \left(\frac{2\kappa_i \theta_{it}}{\xi_{it}^2} \right) \geq 1, \quad \text{for all } i = 1, 2, ..., n.$$

In particular, the last assumption can be considered as the generalization of the Feller condition in the case of time-dependent parameters. Christoffersen et al. (2009) consider a two-factor Heston model in (6.1) with $n = 2$ and constant model parameters. They note that the multifactor Heston model illustrates independence between the level and the slope of volatility smile curves (moneyness[1] effect) and stochastic correlation (term structure effect), by regarding the first factor υ_{1t} and the second factor υ_{2t} as the short-term and long-term variance factors, respectively. As a result, the model is able to provide a rich structure of volatility surfaces that can be observed in the index option market.

6.2.2 Stochastic correlation and the term structure of volatility

We then present some basic properties of the model. In the multifactor Heston model, the instantaneous variances of dX_t and $d\upsilon_t$, and the instantaneous covariance between dX_t and $d\upsilon_t$ are given by

$$\text{Var}[dX_t] = \sum_{i=1}^{n} \upsilon_{it} dt, \quad \text{Var}[d\upsilon_t] = \sum_{i=1}^{n} \xi_i^2 \upsilon_{it} dt, \quad \text{Cov}[dX_t, d\upsilon_t] = \sum_{i=1}^{n} \xi_i \rho_i \upsilon_{it} dt. \tag{6.2}$$

[1] We consider moneyness of a put option under zero interest rate environment in the numerical analysis later. In this case, the moneyness is defined by K/S, where K is the strike price and S is the current stock price.

The level of implied instantaneous volatility is primarily determined by $\text{Var}[dX_t]$, while the volatility skew (i.e. the slope of the implied volatility smile) is determined by $\text{Cov}[dX_t, d\upsilon_t]$. On the other hand, the time-variation of the implied volatility skew can be generated by the stochastic correlation

$$\text{Corr}[dX_t, d\upsilon_t] = \frac{\text{Cov}[dX_t, d\upsilon_t]}{\sqrt{\text{Var}[dX_t]}\sqrt{\text{Var}[d\upsilon_t]}}. \tag{6.3}$$

Under the multifactor Heston model, it is straightforward to compute the expected variance in the case of constant model parameter as

$$\mathbb{E}\left[\upsilon_s \middle| \mathcal{F}_t\right] = \sum_{i=1}^{n} \left[\theta_i \left(1 - e^{-\kappa_i (T-t)}\right) + \upsilon_{i,0} e^{-\kappa_i (T-t)}\right].$$

Hence, the fair strike of a continuously-monitoring variance swap for the contract period $[t, T]$ can be readily obtained as

$$VS\left(t, T\right) = \mathbb{E}\left[\frac{1}{T-t}\int_t^T \upsilon_s ds \middle| \mathcal{F}_t\right] = \sum_{i=1}^{n}\left[\theta_i + \left(\upsilon_{i,0} - \theta_i\right)\frac{\left(1 - e^{-\kappa_i (T-t)}\right)}{\kappa_i \left(T-t\right)}\right]. \tag{6.4}$$

For the pricing of variance swaps under a discrete monitoring, the reader may refer to Zheng and Kwok (2014) and the references therein. As shown in Christoffersen et al. (2009), the multifactor Heston model can generate a rich flexibility for the term structure of implied volatility. Similar multifactor volatility models have been considered in the pricing of volatility derivatives (ee, for example, Bergomi, 2008; Buehler, 2006). It is worth noting that given the model parameters (θ_i, κ_i) for $i = 1, 2, ..., n$, one can explicitly back out the instantaneous variance $\upsilon_{i,0}$ for $i = 1, 2, ..., n$ from n observed market quotes of variance swaps.

6.3 Asymptotic Expansion for Multifactor Heston Model

In this section, we develop an asymptotic expansion for the put option price under the multifactor Heston model with the spirit of Benhamou et al. (2010) who studied the one-factor Heston model.

6.3.1 The perturbed multifactor Heston model

The perturbed processes X_t^ϵ and $\upsilon_{it}^{\epsilon_i}$ of (6.1) are parameterized by $\epsilon = \{\epsilon_i \in [0,1],\, i = 1, 2, ..., n\}$ and are defined as the solution of SDEs

$$\begin{aligned} X_t^\epsilon &= x_0 + \sum_{i=1}^{n} \left[\int_0^t \sqrt{\upsilon_{is}^{\epsilon_i}} dW_s^i - \frac{1}{2} \int_0^t \upsilon_{is}^{\epsilon_i} ds \right], \\ \upsilon_{it}^{\epsilon_i} &= \upsilon_{i0} + \int_0^t \kappa_i(\theta_{is} - \upsilon_{is}^{\epsilon_i}) ds + \epsilon_i \int_0^t \xi_{is} \sqrt{\upsilon_{is}^{\epsilon_i}} dB_t^i. \end{aligned} \tag{6.5}$$

It is clear that when $\epsilon_i = 0$, the perturbed variance process υ_{it}^0 is deterministic

$$\upsilon_{i0,t} \triangleq \upsilon_{it}^0 = e^{-\kappa_i t} \left(\upsilon_{i0} + \int_0^t e^{\kappa_i s} \kappa_i \theta_{is} ds \right),$$

while, when $\epsilon_i = 1$, υ_{it}^1 coincides with υ_{it}, the original variance process. It is noted that under Assumption I, each variance factor $\upsilon_{it}^{\epsilon_i}$ is guaranteed to be positive. Our aim is to obtain an asymptotic expansion of the discounted expectation $g(\epsilon)$ defined by

$$g(\epsilon) = e^{-\int_0^T r_t dt} \mathbb{E} \left[\left(K - e^{\int_0^T (r_t - q_t) dt + X_T^\epsilon} \right)_+ \right], \tag{6.6}$$

where K is the strike price, T is the expiry, r_t is the deterministic risk-free rate and q_t is the deterministic dividend yield. When we set $\epsilon = \mathbf{1}_n$, i.e. $\epsilon_i = 1$ for all $i = 1, 2, ..., n$, the discounted expectation gives the put option pricing formula under the multifactor Heston model.

As the Brownian motions W_t^i are correlated with B_t^i, it is convenient to rewrite (6.5) as

$$X_t^\epsilon = x_0 + \sum_{i=1}^{n} \left[\int_0^t \sqrt{\upsilon_{is}^{\epsilon_i}} \rho_{is} dB_s^i + \int_0^t \sqrt{\upsilon_{is}^{\epsilon_i}} \sqrt{1 - \rho_{it}^2} dZ_s^i \right] - \frac{1}{2} \sum_{i=1}^{n} \int_0^t \upsilon_{is}^{\epsilon_i} ds \tag{6.7}$$

with another set of mutually independent Brownian motions Z_t^i satisfying $d\langle B^i, Z^j \rangle_t = 0$ for all i and j [2]. We denote by $\mathcal{F}_T^B$ the σ-algebra generated by the Brownian motions B_t^i $(i = 1, 2, \ldots, n)$ up to time T. It is easy to see that X_T^ϵ conditional on $\mathcal{F}_T^B$ is Gaussian

[2] An alternative transformation of (6.5) is given by

$$X_t^\epsilon = x_0 + \sum_{i=1}^{n} \int_0^t \sqrt{\upsilon_{is}^{\epsilon_i}} \rho_{is} dB_s^i + \int_0^t \sqrt{\sum_{i=1}^{n} \upsilon_{is}^{\epsilon_i} (1 - \rho_{it}^2)} dZ_s^0 - \frac{1}{2} \sum_{i=1}^{n} \int_0^t \upsilon_{is}^{\epsilon_i} ds,$$

where Z^0 is a Brownian motion which is independent of B^i $(i = 1, 2, \ldots, n)$. This expression may be a more efficient transformation of (6.5) than (6.7) though (6.7) is easier to understand intuitively. For our analysis, the outcoming results will be same whichever is applied.

distributed. Therefore, we can express the discounted expectation as

$$g(\epsilon) = \mathbb{E}\left[e^{-\int_0^T r_t dt}\mathbb{E}\left[\left(K - e^{\int_0^T (r_t - q_t)dt + X_T^\epsilon}\right)_+ \middle| \mathcal{F}_T^B\right]\right] = \mathbb{E}\left[P\left(x\left(\epsilon\right), y\left(\epsilon\right)\right)\right], \tag{6.8}$$

where

$$\begin{aligned} x\left(\epsilon\right) &\triangleq x_0 + \sum_{i=1}^{n}\left[\int_0^T \rho_{it}\sqrt{v_{it}^{\epsilon_i}}dB_t^i - \frac{1}{2}\int_0^T \rho_{it}^2 v_{it}^{\epsilon_i}dt\right], \\ y\left(\epsilon\right) &\triangleq \sum_{i=1}^{n}\int_0^T \left(1-\rho_{it}^2\right)v_{it}^{\epsilon_i}dt, \end{aligned} \tag{6.9}$$

and $P(x, y)$ is the Black-Scholes formula for put option

$$\begin{aligned} &P(x,y) \triangleq Ke^{-r_{eq}T}\mathcal{N}(d) - e^x e^{-q_{eq}T}\mathcal{N}(d-\sqrt{y}), \\ &\text{where} \quad d = \frac{1}{\sqrt{y}}\ln\left[\frac{Ke^{-r_{eq}T}}{e^x e^{-q_{eq}T}}\right] + \frac{1}{2}\sqrt{y},\ r_{eq} = \frac{1}{T}\int_0^T r_t dt,\ q_{eq} = \frac{1}{T}\int_0^T q_t dt, \end{aligned} \tag{6.10}$$

and $\mathcal{N}(\cdot)$ is the standard cumulative normal distribution.

6.3.2 Asymptotic expansion

To expand $g(\epsilon)$ around $\epsilon = \mathbf{0}_n$ asymptotically, let us rewrite (6.8) as

$$g(\epsilon) = \mathbb{E}[P(x' + \Delta x\left(\epsilon\right), y' + \Delta y\left(\epsilon\right))] \tag{6.11}$$

by decomposing (6.9) as $x\left(\epsilon\right) = x' + \Delta x\left(\epsilon\right)$ and $y\left(\epsilon\right) = y' + \Delta y\left(\epsilon\right)$, in which the two terms

$$\begin{aligned} x' &\triangleq x(\mathbf{0}_n) = x_0 + \sum_{i=1}^{n}\left[\int_0^T \rho_{it}\sqrt{v_{i0,t}}dB_t^i - \frac{1}{2}\int_0^T \rho_{it}^2 v_{i0,t}dt\right], \\ y' &\triangleq y(\mathbf{0}_n) = \sum_{i=1}^{n}\int_0^T \left(1-\rho_{it}^2\right)v_{i0,t}dt \end{aligned}$$

represent the values of $X_T^{\mathbf{0}_n}$ conditional on $\mathcal{F}_T^B$ and $\mathrm{Var}(X_T^{\mathbf{0}_n}\big|\,\mathcal{F}_T^B)$, respectively, at the origin of the expansion, $\epsilon = \mathbf{0}_n$, and the rests

$$\begin{aligned} \Delta x\left(\epsilon\right) &= \sum_{i=1}^{n}\Gamma_{iT}^{\epsilon_i}, \text{ with } \Gamma_{iT}^{\epsilon_i} \triangleq \int_0^T \rho_{it}(\sqrt{v_{it}^{\epsilon_i}} - \sqrt{v_{i0,t}})dB_{it} - \int_0^T \frac{1}{2}\rho_{it}^2(v_{it}^{\epsilon_i} - v_{i0,t})dt, \\ \Delta y\left(\epsilon\right) &= \sum_{i=1}^{n}\Xi_{iT}^{\epsilon_i}, \text{ with } \Xi_{iT}^{\epsilon_i} \triangleq \int_0^T (1-\rho_{it}^2)(v_{it}^{\epsilon_i} - v_{i0,t})dt \end{aligned} \tag{6.12}$$

capture the perturbed change of $X_T^\epsilon|\,\mathcal{F}_T^B$ and $\mathrm{Var}(X_T^\epsilon|\,\mathcal{F}_T^B)$, respectively, with respect to ϵ. To proceed, we introduce the multivariate Taylor expansion residual with respect to ϵ for a process $\{Y_t^\epsilon\}$ as

$$R_{\ell,t}^{Y^\epsilon} = R_{\ell,t}\left[Y^\epsilon\right] \triangleq Y_t^\epsilon - \sum_{m=0}^{\ell} \frac{1}{m!} \left(\sum_{i=1}^{n} h_i \frac{\partial}{\partial \epsilon_i} \right)^m Y_t^\epsilon \Big|_{\epsilon=\mathbf{0}_n, \{h_i\}=\{\epsilon_i\}}.$$

When ϵ is a single variate $\epsilon = \epsilon_i \mathbf{1}_i$, the above definition for a process $\{Y_t^{\epsilon_i}\}$ is equivalent to the case of one-factor in Benhamou et al. (2010)

$$R_{\ell,t}^{Y^{\epsilon_i}} = Y_t^{\epsilon_i} - \sum_{m=0}^{\ell} \frac{\epsilon_i^m}{m!} Y_{m,t}^{\epsilon_i}, \quad Y_{m,t}^{\epsilon_i} \triangleq \frac{\partial^m Y_t^{\epsilon_i}}{\partial \epsilon_i^m}\Big|_{\epsilon_i=0}.$$

With these notations, we see that the second order expansion of the variance factors are given by

$$\begin{aligned}
\upsilon_{it}^{\epsilon_i} &= \upsilon_{i0,t} + \epsilon_i \upsilon_{i1,t} + \frac{1}{2!}\epsilon_i^2 \upsilon_{i2,t} + R_{2,t}^{\upsilon_i^{\epsilon_i}}, \\
\sqrt{\upsilon_{it}^{\epsilon_i}} &= \sqrt{\upsilon_{i0,t}} + \epsilon_i \frac{\upsilon_{i1,t}}{2(\upsilon_{i0,t})^{1/2}} + \frac{1}{2!}\epsilon_i^2 \left(\frac{\upsilon_{i2,t}}{2(\upsilon_{i0,t})^{1/2}} - \frac{(\upsilon_{i1,t})^2}{4(\upsilon_{i0,t})^{3/2}} \right) + R_{2,t}^{\sqrt{\upsilon_i^{\epsilon_i}}},
\end{aligned}$$

with

$$\begin{aligned}
&\upsilon_{i0,t} = e^{-\kappa_i t}\left(\upsilon_{i0} + \int_0^t e^{\kappa_i s} \kappa_i \theta_{is} ds \right), \quad \upsilon_{i1,t} = e^{-\kappa_i t} \int_0^t e^{\kappa_i s} \xi_{is} \sqrt{\upsilon_{i0,s}} dB_s^i, \\
&\upsilon_{i2,t} = e^{-\kappa_i t} \int_0^t e^{\kappa_i s} \xi_{is} \frac{\upsilon_{i1,s}}{(\upsilon_{i0,s})^{1/2}} dB_s^i.
\end{aligned} \tag{6.13}$$

As a result, the second order expansion for $\Gamma_{iT}^{\epsilon_i}$ and $\Xi_{iT}^{\epsilon_i}$, defined in (6.12), are given by

$$\Gamma_{iT}^{\epsilon_i} = \Gamma_{i0,T} + \epsilon_i \Gamma_{i1,T} + \frac{1}{2}\epsilon_i^2 \Gamma_{i2,T} + R_{2,T}^{\Gamma_i^{\epsilon_i}}$$

with

$$\begin{aligned}
\Gamma_{i0,T} &= \Gamma_{iT}^0 = 0, \quad \Gamma_{i1,T} = \int_0^T \rho_{it} \frac{\upsilon_{i1,t}}{2(\upsilon_{i0,t})^{1/2}} dB_t^i - \int_0^T \frac{\rho_{it}^2}{2} \upsilon_{i1,t} dt, \\
\Gamma_{i2,T} &= \int_0^T \rho_{it} \left[\frac{\upsilon_{i2,t}}{2(\upsilon_{i0,t})^{1/2}} - \frac{(\upsilon_{i1,t})^2}{4(\upsilon_{i0,t})^{3/2}} \right] dB_t^i - \int_0^T \frac{\rho_{it}^2}{2} \upsilon_{i2,t} dt,
\end{aligned}$$

and

$$\Xi_{iT}^{\epsilon_i} = \Xi_{i0,T} + \epsilon_i \Xi_{i1,T} + \frac{1}{2}\epsilon_i^2 \Xi_{i2,T} + R_{2,T}^{\Xi_i^{\epsilon_i}}$$

with

$$\Xi_{i0,T} = \Xi_{iT}^0 = 0, \quad \Xi_{i1,T} = \int_0^T (1-\rho_{it}^2) \upsilon_{i1,t} dt, \quad \Xi_{i2,T} = \int_0^T (1-\rho_{it}^2) \upsilon_{i2,t} dt.$$

These expansion terms will be used in the following derivation.

Now it is ready to expand (6.11) with respect to ϵ. By the application of the chain rule for the formal derivatives of the parameterized stochastic processes $\Gamma_{iT}^{\epsilon_i}$ and $\Xi_{iT}^{\epsilon_i}$ with respect to $\epsilon = (\epsilon_i)_i$, up to the second order, we arrive at the following expansion formula

$$\begin{aligned} g(\epsilon) &= \mathbb{E}\left[P(x',y')\right] + \mathbb{E}\left[\frac{\partial P(x',y')}{\partial x}\sum_{i=1}^{n}\Delta\Gamma_i^1\right] + \mathbb{E}\left[\frac{\partial P(x',y')}{\partial y}\sum_{i=1}^{n}\Delta\Xi_i^1\right] \\ &\quad + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P(x',y')}{\partial x^2}\left(\sum_{i=1}^{n}\Delta\Gamma_i^2\right)^2\right] + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P(x',y')}{\partial y^2}\left(\sum_{i=1}^{n}\Delta\Xi_i^2\right)^2\right] \\ &\quad + \mathbb{E}\left[\frac{\partial^2 P(x',y')}{\partial x\partial y}\left(\sum_{i=1}^{n}\Delta\Xi_i^2\right)\left(\sum_{i=1}^{n}\Delta\Gamma_i^2\right)\right] + \widetilde{\varepsilon}_n, \end{aligned} \tag{6.14}$$

where

$$\begin{aligned} \Delta\Gamma_i^1 &\triangleq \epsilon_i\Gamma_{i1,T} + \frac{\epsilon_i^2}{2}\Gamma_{i2,T} \\ &= \int_0^T \rho_{it}\left[\epsilon_i\frac{\upsilon_{i1,t}}{2(\upsilon_{i0,t})^{1/2}} + \frac{1}{2!}\epsilon_i^2\left(\frac{\upsilon_{i2,t}}{2(\upsilon_{i0,t})^{1/2}} - \frac{(\upsilon_{i1,t})^2}{4(\upsilon_{i0,t})^{3/2}}\right)\right]dB_t^i \\ &\quad - \int_0^T \frac{1}{2}\rho_{it}^2\left(\epsilon_i\upsilon_{i1,t} + \frac{1}{2}\epsilon_i^2\upsilon_{i2,t}\right)dt, \\ \Delta\Gamma_i^2 &\triangleq \epsilon_i\Gamma_{i1,T} = \int_0^T \rho_{it}\epsilon_i\frac{\upsilon_{i1,t}}{2(\upsilon_{i0,t})^{1/2}}dB_t^i - \int_0^T \frac{1}{2}\rho_{it}^2\epsilon_i\upsilon_{i1,t}dt, \\ \Delta\Xi_i^1 &\triangleq \epsilon_i\Xi_{i1,T} + \frac{\epsilon_i^2}{2}\Xi_{i2,T} = \int_0^T \left(1-\rho_{it}^2\right)\left(\epsilon_i\upsilon_{i1,t} + \frac{1}{2}\epsilon_i^2\upsilon_{i2,t}\right)dt, \\ \Delta\Xi_i^2 &\triangleq \epsilon_i\Xi_{i1,T} = \int_0^T \left(1-\rho_{it}^2\right)\epsilon_i\upsilon_{i1,t}dt, \end{aligned} \tag{6.15}$$

for $i = 1, 2, ..., n$, and $\widetilde{\varepsilon}_n$ is the expansion error. The detailed derivation of (6.14) is found in the Appendix.

By expanding the quadratic terms in the expectations, (6.14) is rewritten as

$$\begin{aligned} g(\epsilon) &= \mathbb{E}\left[P\left(x',y'\right)\right] \\ &\quad + \sum_{i=1}^{n}\Big\{\mathbb{E}\left[\frac{\partial P}{\partial x}\Delta\Gamma_i^1\right] + \mathbb{E}\left[\frac{\partial P}{\partial y}\Delta\Xi_i^1\right] \\ &\quad + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\left(\Delta\Gamma_i^2\right)^2\right] + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\left(\Delta\Xi_i^2\right)^2\right] + \mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\Delta\Gamma_i^2\Delta\Xi_i^2\right]\Big\} \\ &\quad + \sum_{i=2}^{n}\sum_{j=1}^{i-1}\Big\{\mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\Delta\Gamma_i^2\Delta\Gamma_j^2\right] + \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\Delta\Xi_i^2\Delta\Xi_j^2\right] \\ &\quad + \mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\Delta\Gamma_i^2\Delta\Xi_j^2\right] + \mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\Delta\Gamma_j^2\Delta\Xi_i^2\right]\Big\} \\ &\quad + \widetilde{\varepsilon}_n, \end{aligned} \tag{6.16}$$

where the term $\partial^{k+l}P/\partial x^k \partial y^l$ in the expectation is evaluated at (x', y'). The expectations involving the cross terms $\Delta\Gamma_i^2\Delta\Gamma_j^2, \Delta\Xi_i^2\Delta\Xi_j^2$, $\Delta\Gamma_i^2\Delta\Xi_j^2$, and $\Delta\Gamma_j^2\Delta\Xi_i^2$ in (6.16) are new terms that appear in the case of the multifactor Heston model. Although they include the Ito integrals inside the expectation, each of the Ito integral terms can be replaced with some Lebegue-Stieltjes integral thanks to Malliavin calculus. Our interest is the evaluation at $\epsilon = \mathbf{1}_n$. It is summarized in the following Proposition, which is a natural extension of Proposition 2.1 in Benhamou et al. (2010) in the case of multifactor model.

Proposition 6.1. *When $\epsilon = \mathbf{1}_n$, the put option price (6.6) is given by*

$$\begin{aligned} g(\mathbf{1}_n) &= \mathbb{E}\left[P(x', y')\right] + \mathbb{E}\left[\frac{\partial P}{\partial y}(x', y') \sum_{i=1}^{n} \int_0^T (\upsilon_{i1,t} + \upsilon_{i2,t})\, dt\right] \\ &\quad + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x', y') \left(\sum_{i=1}^{n} \int_0^T \upsilon_{i1,t} dt\right)^2\right] + \widetilde{\varepsilon}_n. \end{aligned}$$

Proof. One can show that the summation of terms on the second line and the third one of (6.16) is equal to

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial P}{\partial y}(x', y') \int_0^T (\upsilon_{i1,t} + \upsilon_{i2,t})\, dt\right] + \sum_{i=1}^{n} \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x', y') \left(\int_0^T \upsilon_{i1,t} dt\right)^2\right]$$

in a similar way as in Proposition 2.1 in Benhamou et al. (2010) by applying Lemma 6.11 shown in the Appendix. Similarly, when $\epsilon = \mathbf{1}_n$, the summation inside the parenthesis of the cross terms on the fourth line and the fifth one of (6.16) can be expressed as

$$\begin{aligned} \Phi_T^{i,j} &\triangleq \mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\Delta\Gamma_i^2\Delta\Gamma_j^2\right] + \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\Delta\Xi_i^2\Delta\Xi_j^2\right] + \mathbb{E}\left[\frac{\partial^2 P}{\partial x \partial y}\Delta\Gamma_i^2\Delta\Xi_j^2\right] \\ &\quad + \mathbb{E}\left[\frac{\partial^2 P}{\partial x \partial y}\Delta\Gamma_j^2\Delta\Xi_i^2\right] \\ &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x', y') \int_0^T \left[\int_0^t \upsilon_{i1,s} ds\right] \upsilon_{j1,t} dt\right] \\ &\quad + \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x', y') \int_0^T \left[\int_0^t \upsilon_{j1,s} ds\right] \upsilon_{i1,t} dt\right], \end{aligned} \tag{6.17}$$

whose proof is postponed to the Appendix. Then, by applying the identity

$$\left(\int_0^T f_t dt\right)\left(\int_0^T g_t dt\right) = \int_0^T \left[\int_0^t f_s ds\right] g_t dt + \int_0^T \left[\int_0^t g_s ds\right] f_t dt,$$

we have

$$\Phi_T^{i,j} = \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x', y') \left(\int_0^T \upsilon_{i1,t} dt\right)\left(\int_0^T \upsilon_{j1,t} dt\right)\right]. \tag{6.18}$$

It follows that

$$\begin{aligned} g(\mathbf{1}_n) &= \mathbb{E}\left[P(x',y')\right] + \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial P}{\partial y}(x',y') \int_0^T (\upsilon_{i1,t} + \upsilon_{i2,t})\, dt\right] \\ &\quad + \sum_{i=1}^{n} \frac{1}{2} \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x',y') \left(\int_0^T \upsilon_{i1,t} dt\right)^2\right] \\ &\quad + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x',y') \left(\int_0^T \upsilon_{i1,t} dt\right) \left(\int_0^T \upsilon_{j1,t} dt\right)\right] + \widetilde{\varepsilon}_n \end{aligned} \tag{6.19}$$

which is equivalent to the desired expression. □

An asymptotic expansion formula of the put option price is obtained by the calculation of each expectation in (6.19). For the derivation under the multifactor Heston model, we need the following preliminary results in the one-factor model from Benhamou et al. (2010). In addition, we keep the index i of the single variance factor in the Lemma for later use in the case of the multifactor Heston model.

Lemma 6.2 (The One-factor Model)**.** *Suppose that Assumption I holds and take $n = 1$, $i = 1$, i.e. the one-factor Heston model with time-dependent parameters. Then the put option pricing formula is given by*

$$g(1) = P\left(x_0, var_T^i\right) + P_1^i\left(x_0, var_T^i\right) + \widetilde{\varepsilon}_1,$$

where $var_T^i = \int_0^T \upsilon_{i0,s} dt$ is the total variance,

$$P_1^i(x,y) = \sum_{k=1}^{2} a_{k,T}^i \frac{\partial^{k+1}}{\partial x^k \partial y} P(x,y) + \sum_{k=0}^{1} b_{2k,T}^i \frac{\partial^{2k+2}}{\partial x^{2k} \partial y^2} P(x,y)$$

is the expansion term of the one-factor Heston model when $x = x_0$, $y = var_T^i$, and the expansion coefficients $a_{k,T}^i$ and $b_{2k,T}^i$ are given by

$$\begin{aligned} a_{1,T}^i &= \int_0^T \phi_{i0}(s)\, \phi_{i1}(s)\, ds \int_s^T \phi_{i0}^{-1}(u)\, du, \\ a_{2,T}^i &= \int_0^T \phi_{i0}(s)\, \phi_{i1}(s)\, ds \int_s^T \phi_{i3}(t)\, dt \int_t^T \phi_{i0}^{-1}(u)\, du, \\ b_{0,T}^i &= \int_0^T \phi_{i0}^2(s)\, \phi_{i2}(s)\, ds \int_s^T \phi_{i0}^{-1}(t)\, dt \int_t^T \phi_{i0}^{-1}(u)\, du, \\ b_{2,T}^i &= \frac{1}{2}\left(a_{1,T}^i\right)^2, \end{aligned} \tag{6.20}$$

with

$$\phi_{i0}(s) = e^{\kappa_i s}, \quad \phi_{i1}(s) = \rho_{is}\xi_{is}\upsilon_{i0,s}, \quad \phi_{i2}(s) = \xi_{is}^2 \upsilon_{i0,s}, \quad \phi_{i3}(s) = \rho_{is}\xi_{is}.$$

Proof. See Benhamou et al. (2010). □

In our multifactor setting, there are interacted terms between different variance factors in (6.19)

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x',y')\left(\int_0^T \upsilon_{i1,t}dt\right)\left(\int_0^T \upsilon_{j1,t}dt\right)\right],$$

which leads to additional terms $P_2^{i,j}(x_0, var_T)$ in the approximation formula as indicated in the following Theorem, which is the main contribution in the theoretical part of this work.

Theorem 6.3 (The Multifactor Model)**.** *The put option price (6.6) under the multifactor Heston model is given by*

$$g(\mathbf{1}_n) = P(x_0, var_T) + \sum_{i=1}^{n} P_1^i(x_0, var_T) + \sum_{i=2}^{n}\sum_{j=1}^{i-1} P_2^{i,j}(x_0, var_T) + \widetilde{\varepsilon}_n, \tag{6.21}$$

where $var_T = \sum_{i=1}^{n} var_T^i = \sum_{i=1}^{n}\int_0^T \upsilon_{i0,t}dt$,

$$P_2^{i,j}(x_0, var_T) \;=\; c_T^{i,j}\frac{\partial^4 P}{\partial x^2 y^2}(x_0, var_T), \quad c_T^{i,j} = \mathcal{C}(i,j) + \mathcal{C}(j,i),$$

and

$$\begin{aligned}
\mathcal{B}(i,j,t,T) &= \int_t^T \phi_{i0}^{-1}(u)\left(\int_u^T \phi_{j0}^{-1}(w)dw\right)du + \int_t^T \phi_{j0}^{-1}(u)\left(\int_u^T \phi_{i0}^{-1}(w)dw\right)du,\\
\mathcal{C}(i,j) &= \int_0^T \phi_{i0}(s)\phi_{i1}(s)\left(\int_s^T \phi_{j0}(t)\phi_{j1}(t)\mathcal{B}(i,j,t,T)\,dt\right)ds\\
&+ \int_0^T \phi_{i0}(s)\phi_{i1}(s)\left(\int_s^T \phi_{i0}^{-1}(t)\left(\int_t^T \phi_{j0}(u)\phi_{j1}(u)\left(\int_u^T \phi_{j0}^{-1}(w)dw\right)du\right)dt\right)ds,
\end{aligned} \tag{6.22}$$

with

$$\phi_{i0}(s) = e^{\kappa_i s},\ \ \phi_{i1}(s) = \rho_{is}\xi_{is}\upsilon_{i0,s}.$$

Proof. By following the discussion of proof of Lemma 2, it is straightforward to check that it holds that $P(x_0, var_T) = \mathbb{E}[P(x', y')]$ and

$$\begin{aligned}
&P_1^i(x_0, var_T)\\
=\;& \mathbb{E}\left[\frac{\partial P}{\partial y}(x',y')\int_0^T(\upsilon_{i1,t}+\upsilon_{i2,t})\,dt\right] + \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}(x',y')\left(\int_0^T \upsilon_{i1,t}dt\right)^2\right],
\end{aligned}$$

by making use of Lemma 6.11 and Lemma 6.14 in the Appendix. The sum of these terms corresponds to the first and the second line on the right-hand side of (6.19). Therefore, it remains to calculate the terms (6.18) or (6.17) when $\epsilon = \mathbf{1}_n$. Due to the functional form of the Black-Scholes formula $P(x, y)$, we observe that

$$\mathbb{E}\left[\frac{\partial^{\ell+m}}{\partial x^{\ell}\partial y^{m}} P\left(x', y'\right)\right] = \frac{\partial^{\ell+m}}{\partial x^{\ell}\partial y^{m}} P\left(x_0, var_T\right).$$

Then, as shown in the Appendix, by applying the Malliavin calculus and the Fubini Theorem, the stochastic integrals within the two expectations in (6.17) can be transformed as

$$\Phi_T^{i,j} = c_T^{i,j} \frac{\partial^4 P}{\partial x^2 \partial y^2}\left(x_0, var_T\right), \tag{6.23}$$

which is $P_2^{i,j}\left(x_0, var_T\right)$. □

Theorem 6.3 states that the option price is decomposed into three terms plus an error: (i) the Black-Scholes pricing formula $P\left(x_0, var_T\right)$ with the total variance, (ii) an adjustment term $P_1^i\left(x_0, var_T\right)$ within a specific variance factor, and (iii) an interacting adjustment $P_2^{i,j}\left(x_0, var_T\right)$ by the two different variance factors.

When the parameters are constant, the relevant quantities are calculated explicitly as shown in the following Corollaries.

Corollary 6.4. *When all the model parameters are constant, the variance var_T^i and the expansion coefficients $a_{1,T}^i, a_{2,T}^i$, and $b_{0,T}^i$ in Lemma 6.2 can be explicitly computed as*

$$\begin{aligned} var_T^i &= m_{i0}\upsilon_{i0} + m_1\theta_i, & a_{2,T}^i &= \left(\rho_i\xi_i\right)^2\left(q_{i0}\upsilon_{i0} + q_{i1}\theta_i\right), \\ a_{1,T}^i &= \rho_i\xi_i\left(p_{i0}\upsilon_{i0} + p_{i1}\theta_i\right), & b_{0,T}^i &= \xi_i^2\left(r_{i0}\upsilon_{i0} + r_{i1}\theta_i\right), \end{aligned}$$

with

$$\begin{aligned} m_{i0} &= z_{mi}\left(e^{\kappa_i T} - 1\right), \\ m_{i1} &= T - z_{mi}\left(e^{\kappa_i T} - 1\right), \\ p_{i0} &= z_{pi}\left(-\kappa_i T + e^{\kappa_i T} - 1\right), \\ p_{i1} &= z_{pi}\left(\kappa_i T + e^{\kappa_i T}\left(\kappa_i T - 2\right) + 2\right), \\ q_{i0} &= z_{qi}\left(-\kappa_i T\left(\kappa_i T + 2\right) + 2e^{\kappa_i T} - 2\right), \\ q_{i1} &= z_{qi}\left(2e^{\kappa_i T}\left(\kappa_i T - 3\right) + \kappa_i T\left(\kappa_i T + 4\right) + 6\right), \\ r_{i0} &= z_{ri}\left(-4e^{\kappa_i T}\kappa_i T + 2e^{2\kappa_i T} - 2\right), \\ r_{i1} &= z_{ri}\left(4e^{\kappa_i T}\left(\kappa_i T + 1\right) + e^{2\kappa_i T}\left(2\kappa_i T - 5\right) + 1\right), \end{aligned}$$

and

$$z_{mi} = e^{-\kappa_i T}/(\kappa_i), \quad z_{pi} = e^{-\kappa_i T}/(\kappa_i^2), \quad z_{qi} = e^{-\kappa_i T}/(2\kappa_i^3), \quad z_{ri} = e^{-2\kappa_i T}/(4\kappa_i^3).$$

Corollary 6.5. *When all the model parameters are constant, the expansion coefficients* $c_T^{i,j}$ *in Theorem 6.3 can be explicitly computed as*

$$c_T^{i,j} = \rho_i \rho_j \xi_i \xi_j \left(y_0 \upsilon_{i0} \upsilon_{j0} + y_1 \upsilon_{i0} \theta_j + y_2 \upsilon_{j0} \theta_i + y_3 \theta_i \theta_j \right),$$

where $\kappa(i,k) = \kappa_i T + k, \quad z_y = e^{-(\kappa_i + \kappa_j)T}/(\kappa_i \kappa_j)^2$, *and*

$$\begin{aligned}
y_0 &= z_y \left(e^{\kappa_i T} - \kappa(i,1) \right) \left(e^{\kappa_j T} - \kappa(j,1) \right), \\
y_1 &= z_y \left(e^{\kappa_i T} - \kappa(i,1) \right) \left(e^{\kappa_j T} \kappa(j,-2) + \kappa(j,2) \right), \\
y_2 &= z_y \left[e^{(\kappa_i + \kappa_j)T} \kappa(i,-2) + e^{\kappa_j T} \kappa(j,2) - e^{\kappa_i T} \kappa(i,-2)\kappa(j,1) - \kappa(i,2)\kappa(j,1) \right], \\
y_3 &= z_y \left(e^{\kappa_i T} \kappa(i,-2) + \kappa(i,2) \right) \left(e^{\kappa_j T} \kappa(j,-2) + \kappa(j,2) \right).
\end{aligned}$$

Remark 6.6. As shown in Theorem 6.3, the new expansion terms capture the interaction between different variance factors when the driving Brownian motions W^i and B^i are correlated. In the case of constant parameters, the interaction is related to the covariance as $\rho_i \rho_j \xi_i \xi_j$. In other words, the interaction between two variance factors i and j are induced from its correlation to the underlying process X_t, as given by ρ_i and ρ_j respectively. As the expansion term is linked to $\partial^4 P\left(x_0, var_T\right)/\partial x^2 \partial y^2$, the interaction term is most important for at-the-money-option when the sensitivity in Delta and Vega are significant.

When W^i and B^i are independent, it is observed that $\rho_{it} \equiv 0$ yields $\phi_{i1}(t) \equiv 0$ and $\mathcal{C}(i,j) = 0$. Furthermore, when $\mathcal{C}(i,j) = \mathcal{C}(j,i) = 0$, we see $P_2^{i,j}\left(x_0, var_T\right) = 0$.

Remark 6.7. One may be interested in how the error term $\widetilde{\varepsilon}_n$ behaves as a function of T. Benhamou et al. (2010) shows in Theorem 2.3 that the approximation error for the case of the one-factor Heston model is bounded by $|\widetilde{\varepsilon}_1| \leq C \left(\sup_{t \in [0,T]} \xi_i(t) \right)^3 T^2$ with a generic constant[3] C, which depends on other parameters including T in a nondecreasing way. A similar expression for our multifactor model can be obtained as well. However, the identification of such generic constants as a function of T seems impossible in the derivation process and it is also not so meaningful for the purpose of error estimation, because the generic constants themselves involve T implicitly. Hence, we do not pursue a theoretical evaluation of the error herein. Alternatively, we investigate the errors numerically for several cases in next section.

[3]The precise definition of generic constants is found on p.306 of Benhamou et al. (2010).

Remark 6.8. It is important to note that the approximation formula is derived under the Assumption I. In particular, we impose the Feller condition to be held for all the variance factors. Nevertheless, in practice, when the Heston-type stochastic volatility model is calibrated to market option prices, it is usually found that the Feller condition does not hold. If the Feller condition is imposed during the optimization procedure, the model could give poor fit to the market implied volatility surface. When the Feller condition does not hold, we are not able to guarantee that the approximation formula would produce non-negative option prices which satisfy the no-arbitrage bounds. To resolve this issue, we set the option value to be its intrinsic value when the approximation formula breaches those bounds during the calculation in the calibration section.

6.4 Numerical Illustration

In this section, we study the accuracy of the approximation formula, that is (6.21) by discarding the error term, for the multifactor Heston model. For illustration purposes, we consider the case of $n = 2$, i.e. the case of a two-factor Heston model as in Christoffersen et al. (2009). We consider a couple of scenarios, such as when the model parameters are constant or the correlation coefficients are allowed to be time-dependent.

6.4.1 Constant model parameters

In the multifactor Heston model, the parameter κ_i controls the mean-reversion speed of the i-th variance factor and governs its impact on the term structure of implied volatility. Since the expected variance is given by

$$\mathbb{E}\left[\int_t^T \upsilon_{is} ds \middle| \mathcal{F}_t\right] = \theta_i \left(1 - e^{-\kappa_i (T-t)}\right) + \upsilon_{i,0} e^{-\kappa_i (T-t)},$$

the i-th variance factor could affect the term structure of implied volatility for the time-to-maturity range that is characterized by the half-life (approximately $1/\kappa_i$) of the mean-reversion process. The intuition is that variance factors with different κ_i correspond to the different time-scales of the stochastic volatility process. In the following, we assume that $\kappa_2 > \kappa_1$, such that υ_{1t} can be regarded as a long-term variance factor with a slow mean-reversion, while υ_{2t} can be regarded as a short-term variance factor with a fast mean-reversion.

Accuracy of the approximation formula

To study the accuracy of the approximation formula for the two-factor Heston model, we compute the put option prices in three ways: (i) the characteristic function approach as

the exact price, (*ii*) the approximation formula derived in Section 3, and (*iii*) the Monte-Carlo simulation. For the characteristic function approach, we employ the adaptive integration routines in Matlab (with relative tolerance of 1e-08) to numerically invert the Fourier integral. For the Monte Carlo simulation, we apply the Euler scheme to the simulation of the log-stock price and variance factors, and then adopt the full truncation scheme when the simulated variance path approaches zero. To achieve high convergence, the simulation is repeated 1,000,000 times, while the time-step is kept at 0.01. We take the spot price to be 100 (i.e. $x_0 = \ln(100)$), and consider the range of moneyness between 80% and 120%, with the time-to-maturity of options at 3-, 6-month and 1-, 2-, 3-, 5-, 7-, and 10-year. For simplicity, we assume the interest rate and dividend yield to be zero.

Example 6.1 (Zero Correlation)**.** *We assume the model parameters to be constant as follows:*

ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}
0.0	0.5	0.10	0.25	0.10	0.0	5.0	0.05	0.5	0.05

In Example 6.1, there are no correlations between the stock price and the variance factors. For the model parameters, the first variance factor has a half-life of around 2 years (i.e. $\kappa_1 = 0.5$), while the second variance factor has a half-life of 0.2 year (i.e. $\kappa_2 = 5.0$). We set the mean levels of the long-term and short-term variance factors to be $\theta_1 = 0.10$ and $\theta_2 = 0.05$, respectively (i.e. $\sqrt{0.10} = 31.6\%$ and $\sqrt{0.05} = 22.4\%$ in terms of volatility points, where 1 volatility point = 1%). For simplicity, we set the initial variance (υ_{10} and υ_{20}) of each factor to its corresponding mean level. The Vol-of-Vol for the two variance factors are $\xi_1 = 0.25$ and $\xi_2 = 0.5$. As such, the Feller conditions are given by 1.6 and 2.0, respectively, for the two factors. The parameter setting here is considered to be a moderate market scenario.

In Example 6.1, the approximation formula can be simplified due to zero correlations, resulting in $a^i_{1,T} = a^i_{2,T} = b^i_{2,T} = c^{i,j}_T = 0$. The estimation of the put option prices (as a percentage of the spot price) at various moneyness and time-to-maturity is shown in Table 6.1. We also report the approximation error which is computed by subtracting the exact closed-form price from the approximation price. From Table 6.1, it is found that the approximation formula is very accurate for short-term options, such as the 3-month and 6-month options, in which the approximation errors are between 0 to 0.13 bps (1 bp = 0.0001). The approximation errors for 1-year and 2-year are in the order-of-magnitude of 1 bp and those for longer maturity are, at most, 5 bp. Considering that a typical over-the-counter option is quoted in bp, the approximation formula for short to medium term options are considered to be extremely accurate.

TABLE 6.1: Estimation of put option prices (in percentage of the spot price) for the two-factor Heston model in Example 6.1. MC Error means the standard error of the price by Monte-Carlo simulation.

Moneyness	80	90	100	110	120	80	90	100	110	120
	Time-to-Maturity = 3M					Time-to-Maturity = 6M				
Exact Solution	1.0731	3.3592	7.6739	14.0291	21.9643	2.8353	6.0373	10.8106	17.0469	24.4779
Approximation	1.0731	3.3588	7.6735	14.0286	21.9640	2.8343	6.0363	10.8099	17.0458	24.4765
Approximation Error	0.0000	0.0004	0.0004	0.0005	0.0003	0.0010	0.0010	0.0008	0.0011	0.0013
Monte-Carlo	1.0703	3.3506	7.6660	14.0206	21.9552	2.8380	6.0412	10.8153	17.0543	24.4861
MC Error	0.0035	0.0066	0.0101	0.0133	0.0158	0.0066	0.0101	0.0137	0.0170	0.0199
	Time-to-Maturity =1Y					Time-to-Maturity = 2Y				
Exact Solution	5.9343	9.9852	15.1998	21.4538	28.5809	10.7735	15.6212	21.3036	27.7153	34.7488
Approximation	5.9313	9.9822	15.1970	21.4506	28.5772	10.7629	15.6087	21.2901	27.7014	34.7352
Approximation Error	0.0030	0.0030	0.0029	0.0032	0.0037	0.0106	0.0126	0.0136	0.0139	0.0137
Monte-Carlo	5.9400	9.9992	15.2213	21.4816	28.6140	10.7789	15.6266	21.3099	27.7252	34.7651
MC Error	0.0108	0.0145	0.0181	0.0216	0.0246	0.0160	0.0198	0.0234	0.0269	0.0301
	Time-to-Maturity =3Y					Time-to-Maturity = 5Y				
Exact Solution	14.6067	19.9174	25.9048	32.4787	39.5553	20.7034	26.6173	33.0395	39.8994	47.1362
Approximation	14.5865	19.8936	25.8790	32.4522	39.5294	20.6698	26.5789	32.9981	39.8568	47.0937
Approximation Error	0.0202	0.0238	0.0258	0.0265	0.0259	0.0336	0.0384	0.0413	0.0426	0.0425
Monte-Carlo	14.6404	19.9587	25.9540	32.5337	39.6144	20.7507	26.6734	33.1032	39.9689	47.2094
MC Error	0.0193	0.0231	0.0268	0.0303	0.0335	0.0235	0.0273	0.0309	0.0344	0.0377
	Time-to-Maturity =7Y					Time-to-Maturity = 10Y				
Exact Solution	25.5808	31.9130	38.6593	45.7612	53.1698	31.5315	38.3300	45.4556	52.8616	60.5096
Approximation	25.5426	31.8702	38.6135	45.7137	53.1219	31.4946	38.2893	45.4123	52.8165	60.4636
Approximation Error	0.0382	0.0428	0.0458	0.0474	0.0479	0.0369	0.0407	0.0433	0.0450	0.0460
Monte-Carlo	25.6291	31.9705	38.7222	45.8277	53.2390	31.5624	38.3602	45.4845	52.8892	60.5377
MC Error	0.0260	0.0298	0.0334	0.0369	0.0402	0.0284	0.0321	0.0356	0.0390	0.0423

Example 6.2 (Negative Correlation)**.** *We assume constant model parameters and negative correlations between the stock price and the variance factors with* $\rho_2 < \rho_1 < 0$*.*

ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}
-0.25	0.5	0.10	0.25	0.10	-0.5	5.0	0.05	0.5	0.05

The negative correlation generates the implied volatility skew due to the leverage effect.

In Example 6.2, we allow the correlations between the stock price process and the two variance factors to contrast with those of Example 6.1. As shown in Table 6.2, the approximation errors are generally larger since the underlying dynamics are more complex in the presence of the leverage effect. Nevertheless, the approximation formulae for short-term options up to 1-year remain very accurate with error of less than 1 bp. Meanwhile, in the case of longer-term options, the approximation errors remain well-controlled at the level of, at most, 10 bps.

It is expected that the approximation accuracy could deteriorate sharply in the case of high volatility. To further study the accuracy of the approximation formula, we focus on the 1-year time-horizon, then gradually increase the Vol-of-Vol parameter (ξ_2) for the short-term variance factor (υ_{2t}) at different mean-reversion speeds (κ_2), while keeping other parameters the same as in Example 6.1 (zero correlation).

TABLE 6.2: Estimation of put option prices for the two-factor Heston model in Example 6.2.

Moneyness	80	90	100	110	120	80	90	100	110	120
	Time-to-Maturity = 3M					Time-to-Maturity = 6M				
Exact Solution	1.1831	3.4418	7.6476	13.8943	21.7878	2.9960	6.1177	10.7520	16.8510	24.1929
Approximation	1.1863	3.4422	7.6466	13.8922	21.7836	2.9977	6.1159	10.7500	16.8487	24.1878
Approximation Error	0.0032	0.0003	0.0009	0.0021	0.0042	0.0018	0.0018	0.0020	0.0023	0.0051
Monte-Carlo	1.1770	3.4311	7.6350	13.8831	21.7756	2.9950	6.1159	10.7490	16.8478	24.1876
MC Error	0.0038	0.0069	0.0104	0.0136	0.0161	0.0071	0.0105	0.0141	0.0174	0.0203
	Time-to-Maturity =1Y					Time-to-Maturity = 2Y				
Exact Solution	6.0998	10.0250	15.0789	21.1755	28.1763	10.8630	15.5549	21.0656	27.3108	34.1972
Approximation	6.0960	10.0191	15.0744	21.1722	28.1714	10.8501	15.5381	21.0479	27.2926	34.1776
Approximation Error	0.0039	0.0060	0.0045	0.0033	0.0049	0.0129	0.0168	0.0177	0.0182	0.0196
Monte-Carlo	6.1008	10.0325	15.0925	21.1892	28.1887	10.8590	15.5536	21.0646	27.3100	34.1961
MC Error	0.0114	0.0150	0.0187	0.0221	0.0252	0.0166	0.0203	0.0240	0.0275	0.0308
	Time-to-Maturity =3Y					Time-to-Maturity = 5Y				
Exact Solution	14.5993	19.7455	25.5608	31.9687	38.8944	20.5169	26.2650	32.5213	39.2221	46.3109
Approximation	14.5760	19.7146	25.5250	31.9293	38.8517	20.4733	26.2095	32.4562	39.1492	46.2316
Approximation Error	0.0234	0.0309	0.0358	0.0394	0.0427	0.0436	0.0555	0.0650	0.0728	0.0793
Monte-Carlo	14.6282	19.7775	25.5991	32.0127	38.9424	20.5601	26.3138	32.5763	39.2794	46.3678
MC Error	0.0198	0.0237	0.0274	0.0309	0.0342	0.0240	0.0278	0.0315	0.0350	0.0384
	Time-to-Maturity =7Y					Time-to-Maturity = 10Y				
Exact Solution	25.2502	31.4212	38.0091	44.9595	52.2261	31.0393	37.6852	44.6628	51.9275	59.4421
Approximation	25.1937	31.3520	37.9292	44.8705	52.1293	30.9751	37.6095	44.5770	51.8328	59.3396
Approximation Error	0.0565	0.0692	0.0800	0.0891	0.0968	0.0642	0.0758	0.0859	0.0947	0.1025
Monte-Carlo	25.3074	31.4817	38.0691	45.0177	52.2807	31.0655	37.7093	44.6859	51.9489	59.4623
MC Error	0.0265	0.0303	0.0340	0.0375	0.0409	0.0288	0.0325	0.0361	0.0396	0.0430

Fig. 6.1 plots the approximation errors for each Vol-of-Vol parameter (from 0.1 to 1.0) with a couple of values of moneyness (80%, 100%, 120%) and mean-reversion speed ($\kappa_2 = 5, 4, 3, 2$). As the figure shows, the error increases with the Vol-of-Vol parameter. Additionally, error values for at-the-money options (moneyness of 100%) are, in general, higher than those of other moneyness options. As a higher mean-reversion speed dampens the stochastic movements of the variance process, the approximation error decreases when the mean-reversion speed is high.

Implied volatility surface

Example 6.3. *We show the implied volatility surfaces generated by the following four model parameter settings.*

	ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}
Scenario 1	-0.5	0.5	0.10	0.5	0.05	-0.8	5.0	0.04	1.0	0.02
Scenario 2	-0.5	0.5	0.10	0.5	0.05	-0.8	5.0	0.01	1.0	0.02
Scenario 3	0.0	0.5	0.10	0.5	0.05	-0.8	5.0	0.01	1.0	0.02
Scenario 4	0.5	0.5	0.10	0.5	0.05	-0.8	5.0	0.01	1.0	0.02

Scenario 1 is regarded as the set of baseline parameters. Thence, we change the values of the mean-level for the short-term variance factor (θ_2) and the correlation parameter

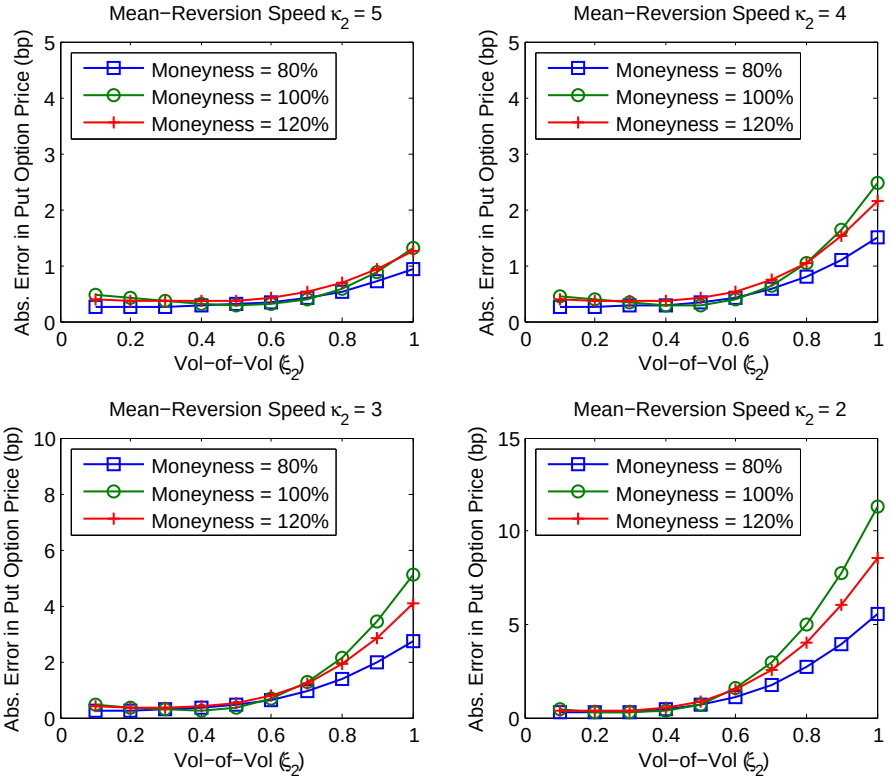


FIGURE 6.1: The plot of the absolute approximation error in put option price with 1-year maturity.

for the long-term factor (ρ_1) in Scenarios 2 to 4. This is in order to produce various shapes of the implied volatility surface. The baseline parameter setting is motivated by the calibration results to the S&P500 index data in Section 5.

- **Scenario 1.** As shown in top-left panel of Fig. 6.2, the implied volatility surface exhibits an moderate upward-sloping term structure because of $\theta_1 > \upsilon_{10}$ and $\theta_2 > \upsilon_{20}$. Given the negative correlation of the two factors ρ_1 and ρ_2, the short-term skewness is prominent and decays gradually with the time-to-maturity. It should be noted that this is the class of implied volatility surface that is commonly observed in the index option market. This is also the case in which a one-factor Heston model is able to provide a good fit.

- **Scenario 2.** To study the impact of the mean levels on the term structure, we change the value of θ_2 from 0.04 to 0.01. This generates a hump-shaped term structure as shown in the top-right panel of Fig. 6.2. The generation of this hump-shaped term structure can be explained based on the variance swap pricing formula (6.4): the instantaneous variance swap rate is given by $\lim_{T\to t} VS(t,T) = \upsilon_{10}+\upsilon_{20}$; if $\tau = T - t$ is $\tau \approx 1/\kappa_2$ but $\tau << 1/\kappa_1$, i.e. in the short-to-medium terms, we have $VS(T,t) \approx \upsilon_{10} + \theta_2$. Hence, the slope of the term structure at the short-end

can be approximated by $VS(t,T) - VS(t,t) \approx \theta_2 - \upsilon_{20}$. In a similar fashion, if $\tau \approx 1/\kappa_1$ and $\tau >> 1/\kappa_2$, the term structure at medium-to-long term can be approximated as $\theta_1 - \upsilon_{10}$. As a result, by tuning the parameters θ_1 and θ_2 relative to υ_{10} and υ_{20}, one can generate a rich variation of the term structure of implied volatility. Moreover, it is worth noting that the change in θ_2 has minimal impact on the 1-month skewness because, in this case, the short-term variance factor takes roughly 3 months to mean-revert.

- **Scenario 3.** Then, we change the long-term correlation ρ_1 from -0.5 to slightly positive at 0.25, while keeping the short-term correlation the same. In this case, the long-term skew flattens out much faster due to the positive leverage effect of the long-term variance factor. This indicates that the long-term skew may be controlled by adjusting the long-term factor. However, the change in ρ_1 also reduces the short-term skewness, indicating that its impact on the short-term smile is not entirely separable.

- **Scenario 4.** Finally, we combine Scenarios 2 and 3 in such a way that a hump-shaped implied volatility surface with a positive skew in long time-to-maturity can be generated. Such a shape of the implied volatility surface is possible when market participants are expecting a medium-term recovery, while perceiving the possibility of a sudden market crash in the near term. Alternatively, it is often observed in the foreign exchange market in which the implied volatility surface is usually more symmetric, with its short-term and long-term smiles to be separately driven by short-term market expectations and long-term macroeconomics factors, respectively.

6.4.2 Time-dependent correlation

In this subsection, we illustrate the modeling of time-dependent correlation under the multifactor Heston model. Given the approximation formula, the put option price under the multifactor Heston model can be easily obtained by a direct numerical integration of (6.20) and (6.22). To compute the iterated integrals, we break down them into nested integrals and apply the trapezoidal rule to convert them into multiple sums. For the characteristic function approach, we numerically solve the system of ordinary differential equations (ODEs) by using the fourth-order Runge-Kutta method and then invert the Fourier transform accordingly. It is worth noting that in the characteristic function approach, for each strike price K, one needs to solve the system of ODEs repeatedly at different grid points during the numerical inversion of the Fourier integral. In contrast, because the expansion coefficients of the approximation formula are independent of the

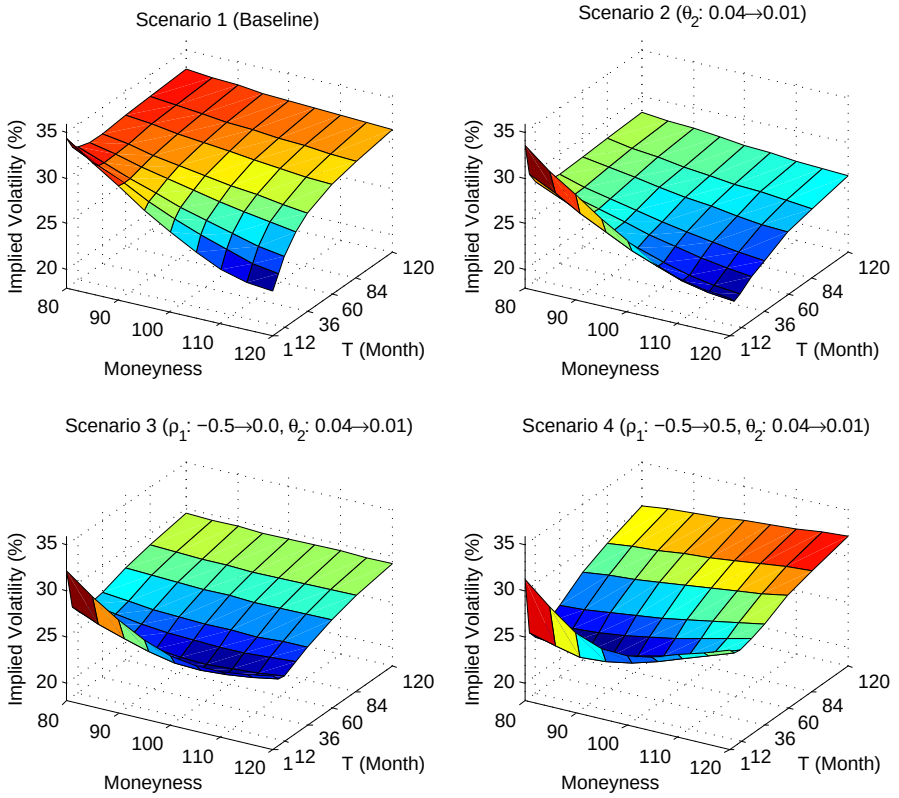


FIGURE 6.2: Implied volatility surfaces in Example 6.3.

strike price, one only needs to compute the expansion coefficients once for a given time-to-maturity to price option at an arbitrary strike.

Example 6.4. *We assume that the correlation of the i-th variance factor with the log stock price starts from $\rho_{i0} = \alpha_i + \chi_i$ and decays exponentially toward χ_i with the deterministic dynamics of*

$$\rho_{it} = \alpha_i e^{-\beta_i t} + \chi_i,$$

where

$$|\rho_{i0}| < 1, \quad \beta_i \geq 0, \quad |\chi_i| \leq 1, \quad |\alpha_i| \leq 1 + |\chi_i|.$$

The parameter β_i governs the convergence speed of the time-dependent correlation to a long-term level of χ_i.

This functional form of the time-dependent correlation is simple and captures the salient fact that the correlation between stock price and volatility decouples with longer time-to-maturity. Alternatively, the specification can be used to produce a persistent long-term skew by setting a high level of χ_i. The time-dependent correlation function allows for greater flexibility in separately modelling the short-term and long-term implied volatility smile.

TABLE 6.3: Estimation of put option prices under the two-factor Heston model in Example 6.4.

Moneyness	80	90	100	110	120	80	90	100	110	120
	Time-to-Maturity = 3M					Time-to-Maturity = 6M				
Numerical ODE	1.1716	3.4341	7.6517	13.9097	21.8071	2.9716	6.1072	10.7644	16.8856	24.2415
Approximation	1.1744	3.4345	7.6514	13.9086	21.8040	2.9730	6.1060	10.7636	16.8848	24.2383
Approximation Error	0.0028	0.0004	0.0003	0.0012	0.0031	0.0014	0.0012	0.0008	0.0008	0.0032
Monte-Carlo	1.1652	3.4232	7.6395	13.9000	21.7962	2.9712	6.1067	10.7633	16.8843	24.2393
MC Error	0.0038	0.0068	0.0104	0.0136	0.0160	0.0070	0.0104	0.0140	0.0174	0.0202
	Time-to-Maturity = 1Y					Time-to-Maturity = 2Y				
Numerical ODE	6.0672	10.0196	15.1082	21.2389	28.2670	10.8410	15.5735	21.1297	27.4193	34.3453
Approximation	6.0647	10.0154	15.1051	21.2363	28.2628	10.8310	15.5592	21.1132	27.4013	34.3253
Approximation Error	0.0024	0.0041	0.0031	0.0026	0.0043	0.0100	0.0143	0.0165	0.0180	0.0200
Monte-Carlo	6.0689	10.0288	15.1251	21.2598	28.2877	10.8395	15.5744	21.1307	27.4210	34.3490
MC Error	0.0113	0.0149	0.0185	0.0220	0.0251	0.0164	0.0202	0.0238	0.0273	0.0306
	Time-to-Maturity = 3Y					Time-to-Maturity = 5Y				
Numerical ODE	14.6021	19.7975	25.6649	32.1236	39.0955	20.5855	26.3939	32.7108	39.4701	46.6136
Approximation	14.5819	19.7698	25.6321	32.0868	39.0558	20.5531	26.3524	32.6621	39.4158	46.5547
Approximation Error	0.0202	0.0277	0.0328	0.0367	0.0397	0.0324	0.0415	0.0487	0.0544	0.0589
Monte-Carlo	14.6315	19.8304	25.7032	32.1695	39.1477	20.6268	26.4441	32.7677	39.5328	46.6793
MC Error	0.0197	0.0235	0.0272	0.0307	0.0340	0.0238	0.0276	0.0313	0.0348	0.0382
	Time-to-Maturity = 7Y					Time-to-Maturity = 10Y				
Numerical ODE	25.3877	31.6251	38.2786	45.2921	52.6180	31.2883	38.0133	45.0702	52.4142	60.0072
Approximation	25.3513	31.5817	38.2296	45.2386	52.5610	31.2429	37.9576	45.0029	52.3331	59.9102
Approximation Error	0.0364	0.0434	0.0490	0.0535	0.0570	0.0454	0.0557	0.0673	0.0811	0.0970
Monte-Carlo	25.4392	31.6823	38.3400	45.3547	52.6807	31.2943	38.0097	45.0544	52.3836	59.9570
MC Error	0.0263	0.0301	0.0338	0.0373	0.0406	0.0286	0.0323	0.0359	0.0394	0.0427

We set the parameters as follows, such that the initial correlation is consistent with Example 6.2 (i.e. $\rho_{10} = -0.25$ and $\rho_{20} = -0.5$).

α_1	β_1	χ_1	α_2	β_2	χ_2
-0.15	0.5	-0.1	-0.25	5.0	-0.25

Table 6.3 shows the accuracy of the approximation formula when correlation coefficients are time-dependent. It is found that the approximation formula remains very accurate even in the case of time-dependent parameters. In fact, the errors are roughly similar to those in Example 6.2. Hence, the approximation formula is still effective in this case.

6.5 Calibration

6.5.1 Data and the calibration procedure

To obtain the risk-neutral model parameters, we perform the daily calibration of the two-factor Heston model using the approximation formula to the cross-sectional market data of index option prices and variance swap rates. We obtain from Bloomberg the interpolated data of implied volatilities of index options for the S&P 500 index (SPX) and Nikkei 225 index (NKY), with fixed maturities of 1-, 2-, 3-, 6-, 12-, 18-, and 24-month

and across moneyness of 80%, 90%, 95%, 97.5%, 100%, 102.5%, 105%, 110%, and 120%. While index options are exchange-traded contracts such that their expiration dates are on fixed calendar days and the strike prices are on standardized grids, Bloomberg interpolates the option implied volatilities and convert into constant maturities and relative moneyness to the spot index level on each trading day. In addition, we obtain variance swap rates for fixed maturities of 1-, 2-, 3-, 6-, 12-, and 24-month on each trading day, which is calculated from the Bloomberg implied volatility using the CBOE VIX index methodology. Hence, the variance swap rates are considered to be theoretical quotes. We also obtain the overnight-index-swap interest rate for the corresponding maturities, which is considered to be a good proxy for the risk-free interest rate in the post-crisis scenario. The dividend yield is assumed to be zero. With these data, we compute the corresponding put option price (referred to as the Bloomberg quoted price) by using the Black-Scholes formula.

As noted in Christoffersen et al. (2009), the calibration of the multifactor Heston model involves the joint-identification of the structural parameters $(\rho_i, \theta_i, \kappa_i, \xi_i)$ and the unobserved initial variance υ_{i0}. They adopt an iterative two-step procedure that separately estimates the structural parameter and the initial variance. Gauthier and Rivaille (2009) note that the initial variance and the mean-reverting level have similar impact on the implied volatility smile. They also suggest that one should avoid the joint-identification of the two parameters during the optimization procedure. Cont and Tankov (2004), meanwhile, discuss the challenges involved in the calibration of an option pricing model to a finite set of market prices as an ill-posed problem. They, in turn, suggest the use of a regularization function to improve the stability of the calibration across different trading days. The slight loss in precision due to the regularization function is justified by the existence of bid-ask spreads, discrete tick in price quote and measurement errors of illiquid options. Against this background, we implement the following two-step procedure to calibrate the model parameters for the two-factor Heston model.

Step 1: Calibration to the term structure of variance swap

As shown in (6.4), the fair strike of variance swap depends only on the structural parameters $\{\theta_1, \theta_2, \kappa_1, \kappa_2\}$ and the two unobserved initial values of variance factor $\{\upsilon_{10}, \upsilon_{20}\}$. Hence, $\Theta_t = \{\theta_1, \theta_2, \kappa_1, \kappa_2, \upsilon_{10}, \upsilon_{20}\}$ are the estimated parameters for Step 1. We calibrate these 6 parameters using the variance swap rates by minimizing the sum-of-square (quadratic) pricing errors as

$$\widehat{\Theta}_t = \arg\min \left[\frac{1}{m} \sum_{k=1}^{m} \left(VS_{k,t}\left(\Theta_t\right) - VS_{k,t} \right)^2 + g\left(\Theta_t\right) \right],$$

where t denotes the trading day, $VS_{k,t}\left(\Theta_t\right)$ and $VS_{k,t}$ are the model-implied and Bloomberg quotes of variance swap rate (in volatility, for example, the fair strike of 20% is taken to be 0.2 in the calibration) for the k-th time-to-maturity respectively. Here, $g\left(\Theta_t\right) = \alpha\left(\Theta_t - \Theta_t^{Int}\right)^2$ is the penalty function that is used to regularize the optimization, where Θ_t^{Int} is the initial guess and α is the loading coefficient of the penalty function (Cont and Tankov, 2004). It is worth noting that the penalty function is incorporated to produce stable estimates for some parameters that are difficult to identify, such as the mean levels $\{\theta_1, \theta_2\}$ and the mean-reversion speed $\{\kappa_1, \kappa_2\}$. We take the initial guess: $\kappa_1 = 0.5$, $\kappa_2 = 5.0, \theta_1 = \upsilon_{10} = 0.10$, and $\theta_2 = \upsilon_{20} = 0.05$. These values are motivated by the average estimates from the trial calibrations which discard the penalty function. From this calibration step, we can identify 6 out of the 10 parameters in the two-factor Heston model.

Step 2: Calibration to option orice and term structure of variance swap

In Step 2, we include the Bloomberg quoted prices of put options and calibrate the two-factor Heston model by minimizing the quadratic pricing error as

$$\begin{aligned}\widehat{\widetilde{\Theta}}_t = \arg\min \quad & \Big[\frac{1}{n}\sum_{j=1}^{n}\omega_j\left(P_{j,t}\left(\widetilde{\Theta}_t\right) - P_{j,t}\right)^2 \\ & + \frac{1}{m}\sum_{k=1}^{m}\left(VS_{k,t}\left(\widetilde{\Theta}_t\right) - VS_{k,t}\right)^2 + g\left(\widetilde{\Theta}_t\right)\Big], \end{aligned} \tag{6.24}$$

where $P_{j,t}\left(\widetilde{\Theta}_t\right)$ is the model-implied price for the j-th put option, and $P_{j,t}$ is the Bloomberg quoted price for the j-th put option. Here, the option price is normalized by the spot price. For the calibration of option price, we select the weighting ω_j to be $(1/Vega_j)^2$, where $Vega_j$ is the Black-Scholes Vega normalized by the spot price as computed using the Bloomberg implied volatility. To avoid giving too much weight to deep in-the-money and out-of-the-money options with very small Vega, we impose a lower bound of $Vega_j$ by 0.01. As noted in Cont and Tankov (2004) and Christoffersen et al. (2009), such a weighting scheme, by using the inverse of Black-Scholes Vega, effectively converts the pricing error in option price into error in implied volatility. In our case, this makes the pricing errors for options and variance swaps to become the same order of magnitude.

Following the suggestion in Gauthier and Rivaille (2009), we exclude the initial values υ_{10} and υ_{20} of variance factors in the calibration in Step 2 and fix them as the estimated values in Step 1. As such, we calibrate the remaining 8 structural parameters $\widetilde{\Theta}_t = \{\rho_1, \theta_1, \kappa_1, \xi_1, \rho_2, \theta_2, \kappa_2, \xi_2\}$ in Step 2. For the initial guess, we take: $\rho_1 = \rho_2 = -0.5$, $\xi_1 = 0.5$ and $\xi_2 = 1.0$. We, then, use the estimated parameter in Step 1 as the initial values for $\theta_1, \theta_2, \kappa_1$, and κ_2. For the optimization algorithm, we use the Levenberg-Marquardt

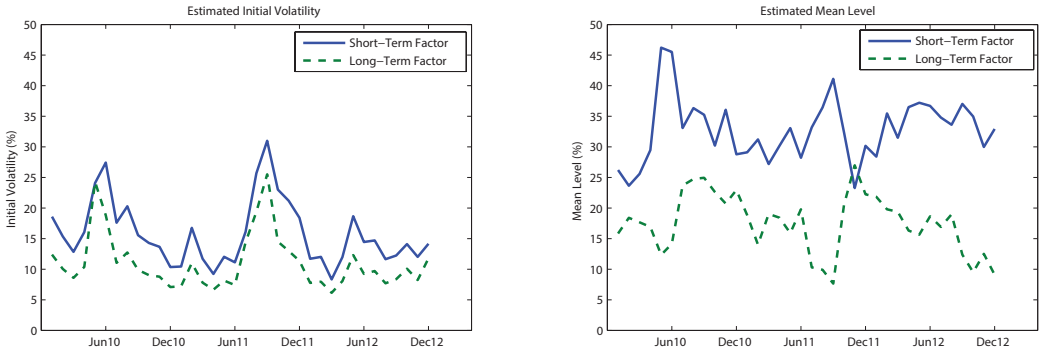


FIGURE 6.3: The time-series dynamics of the estimated initial volatility and long-term mean level in volatility points (computed as square-root of the estimates) from monthly calibration.

algorithm to minimize the quadratic pricing error as a non-linear least-square problem. We carry out the two step calibration once and do not iterate as in Christoffersen et al. (2009).

Remark 6.9. For the penalty function, one has to set a loading coefficient α that balances the stability and precision of the parameter estimates. Cont and Tankov (2004) propose the use of the Morozov discrepancy principle, which authorizes the loss of precision in the optimization procedure that is of the same order of magnitude as the model error when applied to a given data set. In particular, Cont and Tankov (2004) suggest that one can first estimate a priori error level e_0 for the optimization problem (6.24) with $\alpha = 0$. Then, we need to choose the value of α for the penalty function such that the calibration error is slightly better but close to the case when $\alpha = 0$ (i.e., no penalty). Following the procedure, we perform a number of trial calibrations by taking different values of α. Specifically, we take $\alpha = 0.02$ for the SPX market and $\alpha = 0.04$ for the NKY market as reasonable parameters that meet the criteria.

6.5.2 S&P 500 index option

Calibration results

Table 6.4 presents the estimated model parameters from the monthly calibration of the implied volatility surface and the variance swap rates for the sampling period from Jan 2010 to Dec 2012. The instantaneous variance and correlation based on (6.2) and (6.3) are also included as the reference. We calibrate the model at the last Wednesday of each month so as to minimize the month-end liquidity effect that may influence the implied volatility surface. We also report the corresponding calibration errors for the implied volatility surface in terms of volatility points.

TABLE 6.4: Calibrated model parameters and absolute errors (in volatility points) for the SPX market of the two-factor Heston model. The columns Vol. and Corr. are the instantaneous volatility (square-root of the variance) and instantaneous correlation.

	Factor 1					Factor 2							Error		
Date	ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}	Vol.	Corr.	Mean	Min	Max
Year 2010															
Jan	-0.54	0.52	0.069	0.59	0.035	-0.61	5.00	0.025	1.11	0.015	22.3	-0.54	1.27	0.05	3.82
Feb	-0.57	0.46	0.056	0.60	0.024	-0.52	5.00	0.034	0.98	0.010	18.3	-0.53	1.18	0.06	3.21
Mar	-0.61	0.43	0.065	0.56	0.016	-0.52	5.01	0.031	0.86	0.007	15.5	-0.56	1.26	0.07	3.16
Apr	-0.59	0.50	0.087	0.63	0.026	-0.56	5.00	0.029	1.02	0.011	19.1	-0.56	1.49	0.11	3.30
May	-0.55	0.53	0.214	0.77	0.058	-0.80	4.99	0.015	1.45	0.059	34.2	-0.68	1.88	0.03	6.82
Jun	-0.65	0.56	0.207	0.91	0.075	-0.72	4.99	0.020	1.27	0.035	33.3	-0.67	2.16	0.00	5.67
Jul	-0.65	0.44	0.110	0.85	0.031	-0.56	5.00	0.056	1.12	0.012	20.8	-0.61	1.90	0.10	5.14
Aug	-0.73	0.46	0.132	0.83	0.041	-0.61	4.99	0.061	1.09	0.016	23.9	-0.69	1.63	0.00	4.84
Sep	-0.68	0.38	0.124	0.84	0.024	-0.52	5.01	0.062	1.00	0.010	18.4	-0.63	1.78	0.04	5.81
Oct	-0.63	0.37	0.091	0.76	0.020	-0.47	5.01	0.051	0.94	0.008	16.9	-0.58	1.52	0.02	5.08
Nov	-0.63	0.28	0.127	0.69	0.019	-0.51	4.99	0.043	0.94	0.008	16.2	-0.58	1.69	0.00	4.53
Dec	-0.70	0.34	0.083	0.77	0.011	-0.52	5.04	0.053	0.83	0.005	12.6	-0.64	1.79	0.08	5.44
Year 2011															
Jan	-0.64	0.38	0.085	0.66	0.011	-0.50	5.02	0.035	0.81	0.005	12.7	-0.59	1.62	0.05	4.36
Feb	-0.63	0.50	0.097	0.66	0.028	-0.59	5.00	0.020	1.04	0.012	20.1	-0.60	1.71	0.10	3.72
Mar	-0.69	0.38	0.074	0.64	0.014	-0.55	5.02	0.036	0.85	0.006	14.1	-0.63	1.58	0.05	4.53
Apr	-0.66	0.37	0.091	0.65	0.009	-0.49	5.03	0.034	0.74	0.004	11.4	-0.60	1.57	0.11	4.97
May	-0.67	0.43	0.109	0.60	0.014	-0.52	5.01	0.025	0.84	0.007	14.5	-0.60	1.40	0.00	3.96
Jun	-0.67	0.41	0.080	0.65	0.012	-0.53	5.01	0.039	0.82	0.005	13.4	-0.62	1.63	0.10	4.53
Jul	-0.65	0.52	0.110	0.55	0.026	-0.62	5.01	0.011	0.99	0.021	21.6	-0.61	1.22	0.02	4.14
Aug	-0.57	0.56	0.133	0.76	0.066	-0.79	4.99	0.010	1.32	0.038	32.2	-0.65	2.04	0.01	5.67
Sep	-0.63	0.58	0.169	0.83	0.096	-0.93	4.99	0.006	1.47	0.065	40.1	-0.76	2.08	0.01	6.87
Oct	-0.72	0.58	0.105	0.83	0.053	-0.73	4.99	0.042	1.22	0.021	27.2	-0.71	1.90	0.04	4.27
Nov	-0.66	0.50	0.054	0.86	0.045	-0.59	5.00	0.073	1.15	0.017	24.9	-0.63	1.99	0.20	4.62
Dec	-0.67	0.46	0.091	0.74	0.034	-0.57	5.00	0.049	1.05	0.013	21.6	-0.63	1.65	0.02	3.97
Year 2012															
Jan	-0.67	0.39	0.081	0.74	0.014	-0.54	5.02	0.048	0.93	0.006	14.1	-0.62	1.85	0.07	4.92
Feb	-0.67	0.40	0.126	0.73	0.014	-0.50	5.01	0.039	0.90	0.006	14.4	-0.61	1.98	0.12	4.93
Mar	-0.69	0.39	0.099	0.69	0.007	-0.51	5.04	0.038	0.73	0.004	10.3	-0.62	1.80	0.06	4.91
Apr	-0.71	0.44	0.133	0.66	0.014	-0.55	5.01	0.027	0.86	0.007	14.4	-0.64	1.65	0.02	4.22
May	-0.64	0.52	0.138	0.74	0.035	-0.61	4.99	0.025	1.11	0.015	22.4	-0.61	1.93	0.06	4.97
Jun	-0.66	0.39	0.135	0.69	0.021	-0.51	5.00	0.035	0.92	0.009	17.2	-0.60	1.73	0.06	4.49
Jul	-0.64	0.44	0.121	0.65	0.022	-0.51	5.00	0.029	0.93	0.009	17.6	-0.58	1.61	0.01	4.48
Aug	-0.67	0.44	0.113	0.65	0.014	-0.53	5.01	0.036	0.87	0.006	14.0	-0.61	1.53	0.00	4.43
Sep	-0.64	0.45	0.137	0.56	0.015	-0.51	5.01	0.015	0.83	0.007	14.8	-0.58	1.46	0.01	3.82
Oct	-0.63	0.47	0.122	0.54	0.020	-0.52	5.00	0.009	0.88	0.010	17.3	-0.56	1.35	0.02	3.42
Nov	-0.58	0.46	0.090	0.49	0.014	-0.48	5.01	0.016	0.80	0.007	14.6	-0.52	1.22	0.03	3.35
Dec	-0.63	0.50	0.108	0.49	0.020	-0.56	5.01	0.008	0.90	0.014	18.4	-0.56	1.16	0.02	3.64
Summary	ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}	Vol.	Corr.	Mean	Min	Max
Mean	-0.64	0.45	0.110	0.690	0.028	-0.57	5.00	0.033	0.988	0.014	19.3	-0.61	1.645	0.049	4.556
Median	-0.65	0.45	0.109	0.675	0.021	-0.53	5.00	0.034	0.935	0.010	17.5	-0.61	1.641	0.040	4.508
Min	-0.73	0.28	0.054	0.490	0.007	-0.93	4.99	0.006	0.732	0.004	10.3	-0.76	1.162	0.003	3.163
Max	-0.54	0.58	0.214	0.908	0.096	-0.47	5.04	0.073	1.467	0.065	40.1	-0.52	2.165	0.201	6.867

1. **Initial values of variance factors:** The left panel in Fig. 6.3 shows the time series of υ_{10} and υ_{20}, based on the month-end calibration. Given the two-step procedure, the identification of the values of variance factors is very robust with respect to different initial guess and the penalty's loading coefficient. This suggests that the term structure of the variance swap contains rich information about the variance process. The two variance factors can be distinguished as a long-term variance factor with a mean-reversion speed κ_1 of $0.3-0.6$ (i.e. a half-life of around 2 to 3 years), and a short-term variance factor with a mean-reversion speed κ_2 of 5.0 (i.e. a half-life of around 2 to 3 months) as shown in Table 6.4.

2. **Volatility-of-volatility and correlation:** It is well-known that the Vol-of-Vol parameters ξ_1 and ξ_2 capture the level and curvature of the implied volatility surface, while the correlation coefficients ρ_1 and ρ_2 control the skew of the smile. Therefore, the magnitudes of these parameters are expected to be higher during stressed market scenarios. The short-term Vol-of-Vol experiences a sharp increase during the periods of Apr 2010 through Jun 2010, and Jul 2011 through Sep 2011, which correspond to the outbreak of the European debt crisis and the stock market crash amidst the US debt ceiling concerns. In contrast, the long-term Vol-of-Vol remains relatively stable, which reflects the dynamics of a slower timescale. Similar to the dynamics of the Vol-of-Vol parameters, the short-term correlation ρ_2 becomes highly negative at -0.8 and -0.9 during the periods of Apr 2010 through Jun 2010 and Jul 2011 through Sep 2011, respectively. This reflects the steepening of the short-term implied volatility skew when market participants perceive a higher downside risk during a distress market.

3. **Mean level and mean-reversion speed:** As shown in the right panel of Fig. 6.3, the mean levels θ_1 and θ_2 move in tandem (with $\theta_1 > \theta_2$), reflecting the parallel shift of the implied volatility surface with an upward sloping term structure. It is interesting to note that the two parameters move in opposite directions during the periods of Apr 2010 through Jun 2010 and Jul 2011 through Sep 2011, in which the short-term mean level θ_2 experiences a drop, while the long-term mean θ_1 rises. This indicates its freedom to separately control the short-end and long-end of the level of implied volatility surface. Indeed, when we compare the estimated mean levels θ_1 and θ_2 with the variance swap rates of different tenors, we notice that the mean level θ_1 (the long-term factor) is closely linked to the long-term variance swap rate (e.g. the 12-month rate), while the mean level θ_2 appears to control slope of the term structure (e.g. the 12-month rate minus the 1-month rate). On the other hand, the mean-reversion speed parameters κ_1 and κ_2 are very stable across time and are very close to the initial values, when the penalty function is imposed. In fact, the estimates of other model parameters are robust with respect to alternative choices for the initial guess of κ_1 and κ_2. This indicates that the cross-sectional data of option prices at a single trading day do not contain enough information to identify the value of mean-reversion speed. In practice, the mean-reversion speed parameters should be estimated using historical data (e.g. using econometrics techniques) and fixed during the daily calibration exercise.

4. **Calibration errors:** We compute the calibration error, which is the model implied volatility with the calibrated parameters less the Bloomberg implied volatility. As we noted in Remark 6.8, the approximation formula could give an option

value that breaches the no-arbitrage condition for some parameter range and cannot be inverted to the corresponding Black-Scholes implied volatility. Therefore, the model-implied volatility is calculated by using the characteristic function approach, instead of the approximation formula, based on the calibrated parameters. The average calibration error, in terms of implied volatility, ranges from 1% to 2% on different trading days, with the maximum calibration errors ranging from 5% to 7%. It is worth noting that the calibration error is primarily contributed by the pricing error for options in extreme strikes, whereas the average calibration error for around at-the-money options are less than 0.5%, indicating an excellent fit to the implied volatility surface. The poor fit to deep moneyness options can be explained by the model restrictions of the stochastic volatility model and the approximation formula: (*i*) it is known that the short-term skewness at extreme strikes can be best explained by a model with jumps in asset price, such as a jump-diffusion or jump-to-default model; (*ii*) the approximation formula is less accurate for extreme strikes (in percentage terms), making it difficult to fit the implied volatility for these options.

Implied volatility surface

Fig. 6.4 and Fig. 6.5 depict the implied volatility surface of the two consecutive month-end calibration during Apr 2010 and May 2010, which corresponds to the outbreak of the European debt crisis. The calibrated model parameters are as follows.

	ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}
28 Apr 2010	−0.59	0.50	0.087	0.63	0.026	−0.56	5.00	0.029	1.02	0.011
26 May 2010	−0.55	0.53	0.214	0.77	0.058	−0.80	4.99	0.015	1.45	0.059

On 28 Apr 2010 the instantaneous volatility is relatively low (19.1%) while it spiked on 26 May 2010 to 34.1%.

1. **The leverage effect:** Fig. 6.4 shows the calibration result for 28 Apr 2010. The implied volatility surface shows a steep short-term skewness, with the 1-month implied volatility going from 15% at 105% moneyness to 30% at 80% moneyness. As the results show, the two-factor Heston model is able to reproduce the short-term skewness with moderate leverage effect, with the instantaneous correlation of -0.56. In terms of the calibration quality, the pricing errors are overall within 1-2%, with the discrepancy being more significant for short-term and long-term deep in-the-money put options.

2. **The short-term and long-term smile:** Fig. 6.5 shows the calibration result for 26 May 2010. In comparison with the upward-sloping term structure in Fig.

6.4, this latter term structure of implied volatility shows an inverted hump-shape with a significant short-term skewness. This indicates that market participants are expecting a decline of the debt crisis in the medium-term to long-term, while perceiving the possibility of a market crash in the near-term, which can be caused by a sudden change in central bank policy. On the other hand, the short-term skewness decays much faster than in the case of Fig. 6.4. This is because the short-term variance factor υ_{2t} reverts to a lower mean-level θ_2, such that the short-term leverage effect is suppressed. This market scenario corresponds to Fig. 6.3 (right panel), when the mean levels of the two variance factors move in opposite directions.

In terms of the model flexibility, the calibrated implied volatility surface shows that the two-factor model is able to separately control the short-term and long-term term structure and volatility smile during these stress market scenarios. In particular, it should be noted that the two-factor model is able to generate the hump-shaped term structure of variance swap, which is not feasible in the case of the one-factor model. This indicates the necessity to adopt multifactor modeling in order to consistently price European options and volatility derivatives, such as variance swaps.

6.5.3 Nikkei 225 index option

We perform similar monthly calibration using the NKY market data obtained from Bloomberg. In contrast to the SPX option market, the NKY option market is less liquid in deep out-of-the money and long-maturity options (Fukasawa et al., 2011). Thus, we note that the poor liquidity of long-term options in the NKY market may render the interpolation procedure by Bloomberg unreliable. Therefore, one should be cautious in interpreting the calibrated results in these cases. Fortunately, in the presence of the regularization procedure, the calibrated estimates are not very sensitive to the outliers. As a result, we include only the options with moneyness 90%, 95%, 97.5%, 100%, 102.5%, 105%, and 110% in the calibration. The parameter settings and procedures are similar to the case of the SPX options.

Calibration results

The monthly calibration result for the NKY data is presented in Table 6.5. As the implied volatility surface is more flat, the estimate of the short-term mean level θ_1 is found to be small, reflecting that the term structure at the short-term is usually inverted or moderately upward. In comparison with the SPX market, the estimated correlations are negative but closer to zero (with the average instantaneous correlation of -0.48),

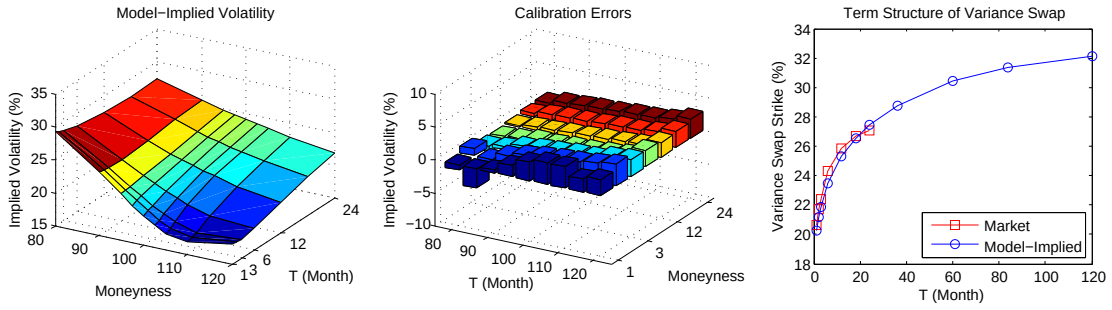


FIGURE 6.4: SPX Index on 28 Apr 2010: the plot of the calibrated model-implied volatility surface, calibration errors, and calibration result to the Bloomberg theoretical term structure of variance swap.

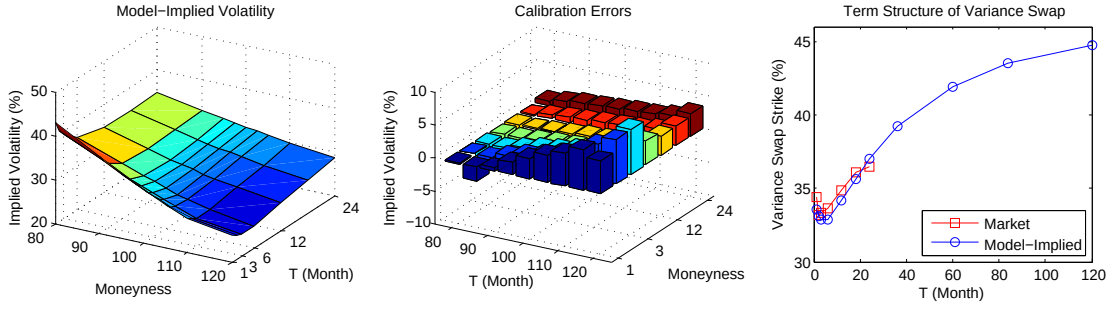


FIGURE 6.5: SPX Index on 26 May 2010: the plot of the calibrated model-implied volatility surface, calibration errors, and calibration result to the Bloomberg theoretical term structure of variance swap.

suggesting the NKY implied volatility surface is more flat. On the other hand, the time-series variations of other model parameters are similar to those of the SPX calibration.

Implied volatility surface

Fig. 6.6 shows the calibrated implied volatility surface on 30 Mar 2011, which captures the stress market after the 2011 Tohoku earthquake. The calibrated model parameters are as follows.

	ρ_1	κ_1	θ_1	ξ_1	υ_{10}	ρ_2	κ_2	θ_2	ξ_2	υ_{20}
30 Mar 2011	−0.52	0.54	0.100	0.47	0.041	−0.70	5.01	0.001	1.21	0.072

The overall fit of the two-factor model for around at-the-money and medium time-to-maturity options are excellent. It is able to capture the inverted term structure and some of the skew of implied volatility. Nevertheless, the model has a difficulty in reproducing

the short-term skew of deep moneyness options, which may be captured more suitably by a jump-diffusion model, given its disaster nature.

TABLE 6.5: Calibrated model parameters and absolute errors (in volatility points) for the NKY market of the two-factor Heston model.

	Factor 1					Factor 2							Error		
Date	ρ_1	κ_1	θ_1	ξ_1	v_{10}	ρ_2	κ_2	θ_2	ξ_2	v_{20}	Vol.	Corr.	Mean	Min	Max
Year 2011															
Jan	-0.48	0.51	0.054	0.48	0.016	-0.50	5.00	0.040	1.01	0.006	14.6	-0.45	0.93	0.02	4.07
Feb	-0.46	0.52	0.066	0.49	0.030	-0.50	5.00	0.016	1.03	0.013	20.6	-0.45	1.18	0.02	4.58
Mar	-0.52	0.54	0.100	0.47	0.041	-0.70	5.01	0.001	1.21	0.072	33.6	-0.62	1.96	0.14	6.86
Apr	-0.56	0.53	0.093	0.56	0.038	-0.58	5.00	0.014	1.08	0.016	23.1	-0.54	1.33	0.00	4.24
May	-0.57	0.55	0.105	0.52	0.038	-0.62	5.00	0.001	1.11	0.018	23.6	-0.55	1.61	0.00	5.15
Jun	-0.48	0.53	0.104	0.58	0.040	-0.56	5.00	0.002	1.08	0.018	24.1	-0.49	1.76	0.10	4.17
Jul	-0.40	0.50	0.060	0.53	0.030	-0.46	5.00	0.024	1.02	0.012	20.7	-0.41	1.59	0.00	4.73
Aug	-0.72	0.57	0.141	0.81	0.063	-0.60	5.00	0.001	1.11	0.023	29.3	-0.67	2.57	0.44	9.74
Sep	-0.60	0.49	0.116	0.80	0.079	-0.62	5.01	0.006	1.16	0.048	35.8	-0.60	3.40	0.10	6.89
Oct	-0.57	0.48	0.135	0.75	0.049	-0.54	5.00	0.022	1.06	0.021	26.5	-0.55	1.89	0.27	4.53
Nov	-0.59	0.53	0.154	0.71	0.063	-0.60	5.00	0.005	1.11	0.032	30.8	-0.58	2.26	0.13	5.72
Dec	-0.47	0.55	0.036	0.45	0.030	-0.53	4.99	0.043	1.05	0.009	19.7	-0.46	1.84	0.08	7.94
Year 2012															
Jan	-0.43	0.51	0.058	0.50	0.029	-0.47	5.00	0.027	1.03	0.012	20.2	-0.42	1.17	0.13	3.45
Feb	-0.44	0.43	0.180	0.85	0.027	-0.43	5.00	0.019	1.00	0.011	19.6	-0.44	2.28	0.28	5.63
Mar	-0.44	0.46	0.069	0.64	0.023	-0.44	5.00	0.031	1.01	0.009	17.8	-0.43	1.50	0.01	4.31
Apr	-0.48	0.47	0.057	0.63	0.025	-0.48	5.00	0.033	1.03	0.010	18.8	-0.47	1.30	0.14	2.92
May	-0.35	0.41	0.227	0.99	0.037	-0.43	5.00	0.016	1.01	0.017	23.3	-0.38	3.30	0.16	8.26
Jun	-0.55	0.44	0.100	0.80	0.025	-0.48	5.00	0.037	1.04	0.010	18.5	-0.52	2.14	0.17	4.43
Jul	-0.51	0.48	0.101	0.72	0.031	-0.52	4.99	0.021	1.06	0.012	20.9	-0.51	2.01	0.07	4.49
Aug	-0.37	0.48	0.108	0.65	0.021	-0.41	4.99	0.022	1.01	0.008	17.1	-0.38	1.67	0.04	4.05
Sep	-0.37	0.45	0.133	0.76	0.019	-0.37	4.99	0.021	0.99	0.008	16.5	-0.37	2.32	0.31	6.41
Oct	-0.40	0.44	0.134	0.67	0.020	-0.44	5.00	0.011	1.00	0.012	17.8	-0.41	1.70	0.00	4.77
Nov	-0.32	0.48	0.058	0.50	0.021	-0.37	5.00	0.019	0.96	0.009	17.2	-0.32	1.67	0.05	3.99
Dec	-0.48	0.48	0.101	0.63	0.029	-0.49	5.00	0.021	1.02	0.012	20.2	-0.47	1.34	0.02	4.18
Summary	ρ_1	κ_1	θ_1	ξ_1	v_{10}	ρ_2	κ_2	θ_2	ξ_2	v_{20}	Vol.	Corr.	Mean	Min	Max
Mean	-0.48	0.49	0.104	0.645	0.034	-0.51	5.00	0.019	1.051	0.017	22.1	-0.48	1.865	0.112	5.230
Median	-0.48	0.48	0.101	0.638	0.030	-0.49	5.00	0.020	1.029	0.012	20.4	-0.46	1.728	0.089	4.558
Min	-0.72	0.41	0.036	0.451	0.016	-0.70	4.99	0.001	0.963	0.006	14.6	-0.67	0.931	0.000	2.915
Max	-0.32	0.57	0.227	0.991	0.079	-0.37	5.01	0.043	1.211	0.072	35.8	-0.32	3.396	0.439	9.739

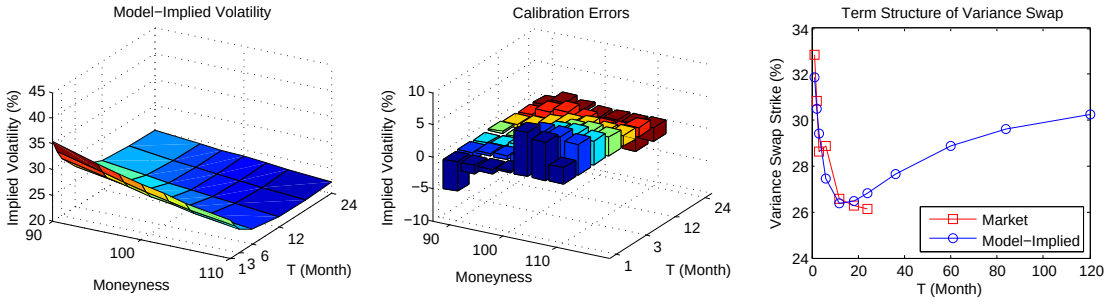


FIGURE 6.6: NKY Index on 30 Mar 2011: the plot of the calibrated model-implied volatility surface, calibration errors, and calibration result to the Bloomberg theoretical term structure of variance swap.

6.5.4 Model-implied long-dated volatility

So far we have calibrated the model parameters based on the SPX and NKY market data up to maturity of 2 years. However, one may be interested in the model-implied volatilities for maturities of longer than 2 years. Fig. 6.7 and 6.8 depict the model-implied volatility surface and the model-implied term structure of variance swap rates, respectively, up to 10 years by using the two-factor Heston model and the calibrated results for the SPX in Section 5.2.2. and the NKY in Section 5.3.2. Each of the resulting term structures seems reasonable although there is no way to verify the accuracy due to the current lack of actual markets.

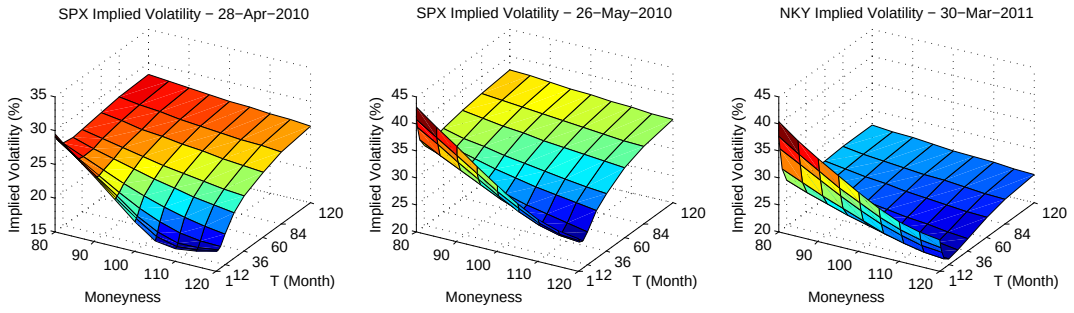


FIGURE 6.7: Model-implied volatility surface up to 10 years by using the calibrated results for the SPX in Section 5.2.2. and the NKY in Section 5.3.2.

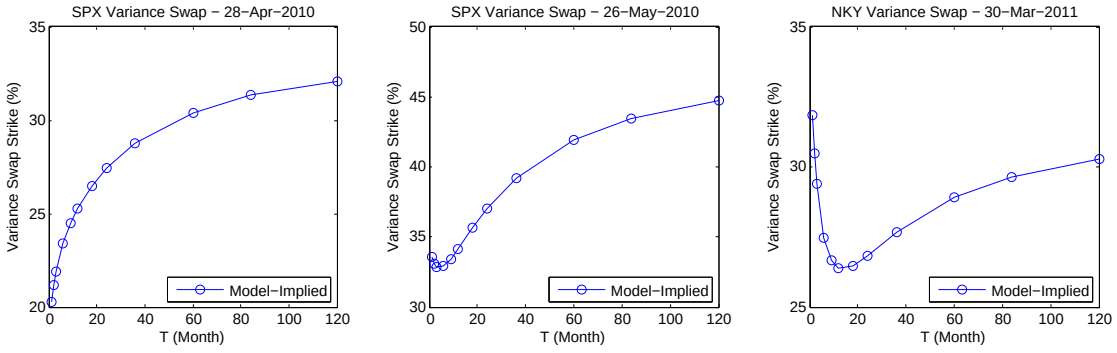


FIGURE 6.8: Model-implied term structure of variance swap rates up to 10 years by using the calibrated results for the SPX in Section 5.2.2. and the NKY in Section 5.3.2.

6.5.5 Computational time

In terms of computational time, the calibration using the approximation formula is very efficient. For example, when we perform the calibration using the Matlab routine for

Levenberg-Marquardt algorithm (running on a laptop with an Intel(R) Core(TM) i7-3520 CPU at 2.90 GHz), and set the convergence tolerance of the objective function to 1e-05, the total computational time to calibrate the 36 snapshots of end-of-month implied volatility surface is around 100-150 seconds. In contrast, when the characteristic function is used to compute the exact option price, the corresponding computational time is 2000 to 2500 seconds. Thus, the computational speed improves significantly by a factor of 20. Indeed, the calibration to a snapshot of the implied volatility surface is almost instantaneous when the approximation formula is used. In addition, we find that the calibration using the approximation formula gives more stable estimated parameters across time. This can be explained by the fact that the computation of the exact option price using the characteristic function approach may encounter numerical instability during the calibration process at different parameter ranges (e.g. the selected contour for Fourier inversion may not be suitable for some extreme parameter ranges). In practice, it is difficult for the researcher to ensure the numerical stability of the Fourier inversion at every single iteration of the optimization procedure. The ease of implementation of the approximation formula suggests that the calibration can be done and visualized even on an excel spreadsheet environment. Moreover, given the gain in computation efficiency, the approximation formula is useful for econometrics estimation, back-testing of the model, as well as evaluation of portfolio risk (e.g. calculation of Value-at-Risk or counterparty exposure), in which one has to evaluate a large number of option prices while precision is of less concern.

6.6 Summary

In this chapter, we develop an asymptotic approach to the multifactor Heston option pricing model under time-dependent model parameters. The expansion terms under constant parameters are explicitly computed, while the incorporation of time-dependent parameters can be achieved in a straightforward manner. For illustration, we calibrate a two-factor Heston model to the option price and term structure of the variance swap of the S&P 500 index. The calibration results show that it is possible to distinguish a short-term and a long-term variance factor with different mean-reversion speeds and levels. In particular, the two-factor model provides the flexibility to separately control the short-end and long-end of the implied volatility, in order to fit to various shapes of the implied volatility surface during stress market scenarios. Moreover, in terms of computational time, the approximation formula speeds up the calibration procedure by at least a factor of 20 relative to the case when the characteristic function approach is used to compute the model prices. As the approximation formula allows one to compute option prices under the Heston model with multifactor extension and time-dependent

parameters in a unified framework, it would be interesting to perform an empirical study that compares the goodness-of-fit of these extensions. Finally, it is worth noting that the asymptotic approach developed in this chapter can be readily applied to other multifactor models, such as a jump-to-default model with stochastic default intensity, or a mixture stochastic volatility model, in which one factor is driven by a Brownian motion and another factor is driven by a fractional Brownian motion. These extensions are left for further research.

6.7 Appendix

6.7.1 Preliminary results with Malliavin calculus

We set out the following definitions and lemmas by following Benhamou et al. (2010) for proofs of some formula in what follows.

Definition 6.10. (Integral Operator)

1. For any real number k and any integrable function l on $[0,T]$, we denote

$$\omega_{t,T}^{(k,l)} \triangleq \int_t^T e^{ku} l(u)\, du,$$

for $t \in [0,T]$.

2. For any real numbers $(k_1, k_2, ..., k_n)$ and any integrable functions $(l_1, l_2, ..., l_n)$ on $[0,T]$, we denote the n-times iterated integral as

$$\omega_{t,T}^{(k_1,l_1),(k_2,l_2),...,(k_n,l_n)} \triangleq \omega_{t,T}^{\left(k_1,l_1\omega_{.,T}^{(k_2,l_2),...,(k_n,l_n)}\right)},$$

for $t \in [0,T]$.

Lemma 6.11. *Suppose that a random variable is given in a form of* $G(V_T) \in \mathbb{D}^{1,\infty}(\Omega)$, *where* G *is a smooth function,* $V_T = \sum_{i=1}^{n} \int_0^T \rho_{it} \sqrt{\upsilon_{i0,t}} dB_t^i$, *and* B^i *are independent standard Brownian motions for* $i = 1, 2, ..., n$. *Let* γ *be a square integrable and predictable process. Then we have*

$$\mathbb{E}\left[G(V_T) \int_0^t \gamma_s dB_s^\alpha\right] = \mathbb{E}\left[G^{(1)}(V_T) \int_0^t \gamma_s \rho_{\alpha t} \sqrt{\upsilon_{\alpha 0,t}} ds\right],$$

for all $\alpha = 1, 2, ..., n$, *where* $G^{(k)}$ *is the* k*-th derivative of* G.

Proof. See Proposition 1.2.3 in Nualart (2006). □

Lemma 6.12. *For any deterministic integrable function f on $[0,T]$ and any continuous semimartingale Z with $Z(0)=0$, we have*

$$\int_0^T f(t)\,Z(t)\,dt = \int_0^T \omega_{t,T}^{(0,f)} dZ(t).$$

Proof. See Benhamou et al. (2010). □

Lemma 6.13. *Let $P = P(x,y)$ to be the Black-Scholes formula (6.10) of a put option. Then it holds that*

$$\left[\frac{\partial}{\partial y} - \frac{1}{2}\left[\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}\right]\right] P(x,y) = 0,$$

for all $x \in \mathbb{R}$ and $y \in \mathbb{R}^+$.

Proof. This can be proved by direct differentiation of the Black-Scholes formula. □

The following Lemma is an extension of Lemma 5.5 in Benhamou et al. (2010).

Lemma 6.14. *Let $G(V_T) = G\left(\sum_{i=1}^n \int_0^T \rho_{it}\sqrt{\upsilon_{i0,t}}dB_t^i\right) \in \mathbb{D}^{1,\infty}(\Omega)$ as in Lemma 6.11, h be a deterministic function which is integrable, and $\upsilon_{i1,t} = e^{-\kappa_i t}\int_0^t e^{\kappa_i s}\xi_{is}\sqrt{\upsilon_{i0,s}}dB_s^i$ as defined in (6.13) for $i = 1,2,...,n$. Then, we have*

$$\begin{aligned}
\mathbb{E}\left[G(V_T)\int_0^T h(t)\,\upsilon_{\alpha 1,t}dt\right] &= \omega_{0,T}^{(\kappa_\alpha,\phi_{\alpha 1}),(-\kappa_\alpha,h)}\mathbb{E}\left[G^{(1)}(V_T)\right], \\
\mathbb{E}\left[G(V_T)\int_0^T h(t)\,\upsilon_{\alpha 1,t}\upsilon_{\beta 1,t}dt\right] &= \omega_{0,T}^{(\kappa_\alpha,\phi_{\alpha 1}),(\kappa_\beta,\phi_{\beta 1}),(-(\kappa_\alpha+\kappa_\beta),h)}\mathbb{E}\left[G^{(2)}(V_T)\right] \\
&\quad + \omega_{0,T}^{(\kappa_\beta,\phi_{\beta 1}),(\kappa_\alpha,\phi_{\alpha 1}),(-(\kappa_\alpha+\kappa_\beta),h)}\mathbb{E}\left[G^{(2)}(V_T)\right], \\
&\quad \alpha \neq \beta,
\end{aligned}$$

where

$$\phi_{\alpha 1}(s) = \rho_{\alpha s}\xi_{\alpha s}\upsilon_{\alpha 0,s}, \quad \phi_{\beta 1}(s) = \rho_{\beta s}\xi_{\beta s}\upsilon_{\beta 0,s}.$$

Proof. The proof is similar to Lemma 5.5 in Benhamou et al. (2010) by using Lemma 6.11 and hence is omitted. □

6.7.2 Proof of (6.14)

Recall that the put option price under the perturbed Heston model is given by

$$g(\epsilon) = \mathbb{E}\left[e^{-\int_0^T r_t dt}\mathbb{E}\left[\left(K - e^{\int_0^T (r_t - q_t)dt + X_T^\epsilon}\right)_+ \middle| \mathcal{F}_T^B\right]\right] = \mathbb{E}[P(x' + \Delta x(\epsilon), y' + \Delta y(\epsilon))].$$

Now, we expand $P(x' + \Delta x(\epsilon), y' + \Delta y(\epsilon))$ with respect to ϵ around $\epsilon = \mathbf{0}_n$ up to the second order,

$$\begin{aligned} P(x' + \Delta x(\epsilon), y' + \Delta y(\epsilon)) &= P(x', y') + \sum_{i=1}^{n} \epsilon_i \left.\frac{\partial P}{\partial \epsilon_i}(x(\epsilon), y(\epsilon))\right|_{\epsilon=\mathbf{0}_n} \\ &\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \epsilon_i\epsilon_j \left.\frac{\partial^2 P}{\partial \epsilon_i \partial \epsilon_j}(x(\epsilon), y(\epsilon))\right|_{\epsilon=\mathbf{0}_n} + \varepsilon, \end{aligned} \tag{6.25}$$

where ε is the expansion error. The partial derivatives are given by chain rules as

$$\begin{aligned} \frac{\partial P}{\partial \epsilon_i}(x(\epsilon), y(\epsilon)) &= P_x \partial_i x + P_y \partial_i y, \\ \frac{\partial^2 P}{\partial \epsilon_i^2}(x(\epsilon), y(\epsilon)) &= P_{xx}(\partial_i x)^2 + P_x \partial_i^2 x + P_{yy}(\partial_i y)^2 + P_y \partial_i^2 y + 2P_{xy}(\partial_i x)(\partial_i y), \\ \frac{\partial^2 P}{\partial \epsilon_i \partial \epsilon_j}(x(\epsilon), y(\epsilon)) &= P_{xx}(\partial_i x)(\partial_j x) + P_{xy}(\partial_i x)(\partial_j y) + P_{yy}(\partial_i y)(\partial_j y) + P_{xy}(\partial_j x)(\partial_i y), \end{aligned}$$

with the notation $P_x = \frac{\partial P}{\partial x}(x', y')$, $\partial_i x = \left.\frac{\partial x(\epsilon)}{\partial \epsilon_i}\right|_{\epsilon=\mathbf{0}_n}$ and $\partial_i \partial_j x = \left.\frac{\partial^2 x(\epsilon)}{\partial \epsilon_i \partial \epsilon_j}\right|_{\epsilon=\mathbf{0}_n}$, where we have used the relationship $\partial_i \partial_j x(\epsilon) = \partial_i \partial_j y(\epsilon) = 0$ due to the form of (6.9). Substituting these derivatives in the expansion formula (6.25), we have

$$\begin{aligned} &P(x' + \Delta x(\epsilon), y' + \Delta y(\epsilon)) \\ &= P(x', y') + \sum_{i=1}^{n}\left[P_x\left(\epsilon_i \partial_i x + \frac{1}{2}\epsilon_i^2 \partial_i^2 x\right) + P_y\left(\epsilon_i \partial_i y + \frac{1}{2}\epsilon_i^2 \partial_i^2 y\right)\right] \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \epsilon_i\epsilon_j\left[P_{xx}(\partial_i x)(\partial_j x) + P_{yy}(\partial_i y)(\partial_j y) + 2P_{xy}(\partial_i x)(\partial_j y)\right] + \varepsilon. \end{aligned}$$

By noting that

$$\begin{aligned} \partial_i x(\epsilon) &= \partial_i \Gamma_{iT}^{\epsilon_i} = \Gamma_{i1,T}, \quad \partial_i^2 x(\epsilon) = \partial_i^2 \Gamma_{iT}^{\epsilon_i} = \Gamma_{i2,T}, \\ \partial_i y(\epsilon) &= \partial_i \Xi_{iT}^{\epsilon_i} = \Xi_{i1,T}, \quad \partial_i^2 y(\epsilon) = \partial_i^2 \Xi_{iT}^{\epsilon_i} = \Xi_{i2,T}, \end{aligned}$$

the expansion is written as

$$\begin{aligned}
& P(x'+\Delta x(\epsilon), y'+\Delta y(\epsilon)) \\
= \; & P(x',y') + \sum_{i=1}^{n}\left[P_x\left(\epsilon_i\Gamma_{i1,T}+\frac{\epsilon_i^2}{2}\Gamma_{i2,T}\right)+P_y\left(\epsilon_i\Xi_{i1,T}+\frac{\epsilon_i^2}{2}\Xi_{i2,T}\right)\right] \\
& +\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\epsilon_i\epsilon_j\left[P_{xx}\Gamma_{i1,T}\Gamma_{j1,T}+P_{yy}\Xi_{i1,T}\Xi_{j1,T}+2P_{xy}\Gamma_{i1,T}\Xi_{j1,T}\right]+\varepsilon, \\
= \; & P(x',y') + P_x\sum_{i=1}^{n}\left(\epsilon_i\Gamma_{i1,T}+\frac{\epsilon_i^2}{2}\Gamma_{i2,T}\right)+P_y\sum_{i=1}^{n}\left(\epsilon_i\Xi_{i1,T}+\frac{\epsilon_i^2}{2}\Xi_{i2,T}\right) \\
& +\frac{1}{2}\left[P_{xx}\left(\sum_{i=1}^{n}\epsilon_i\Gamma_{i1,T}\right)^2+P_{yy}\left(\sum_{i=1}^{n}\epsilon_i\Xi_{i1,T}\right)^2+2P_{xy}\left(\sum_{i=1}^{n}\epsilon_i\Gamma_{i1,T}\right)\left(\sum_{i=1}^{n}\epsilon_i\Xi_{i1,T}\right)\right]+\varepsilon.
\end{aligned}$$

Then, by taking the expectations on the both sides and denoting $\tilde{\varepsilon}=\mathbb{E}[\varepsilon]$, we obtain the result.

6.7.3 Proof of (6.17)

From the expansion formula, it is easy to verify that when $\epsilon=\mathbf{1}_n$, take $i=\alpha$ and $j=\beta$ such that $\alpha\neq\beta$, we have

$$\Phi_T^{\alpha,\beta}=\mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}H_{\alpha T}H_{\beta T}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}L_{\alpha T}L_{\beta T}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}H_{\alpha T}L_{\beta T}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}H_{\beta T}L_{\alpha T}\right],$$

where

$$H_{\alpha t}=\int_0^t\rho_{\alpha s}\frac{\upsilon_{\alpha 1,s}}{2(\upsilon_{\alpha 0,s})^{1/2}}dB_s^{\alpha}-\frac{1}{2}\int_0^t\rho_{\alpha s}^2\upsilon_{\alpha 1,s}ds,\quad L_{\alpha t}=\int_0^T\left(1-\rho_{\alpha s}^2\right)\upsilon_{\alpha 1,s}ds.$$

By the application of Ito's Lemma and the independence of $\left\{B_t^i:i=1,2,...,n\right\}$, we can express

$$\begin{aligned}
\Phi_T^{\alpha,\beta} \; = \; & \mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\int_0^T H_{\alpha t}dH_{\beta t}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\int_0^T H_{\beta t}dH_{\alpha t}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T L_{\alpha t}dL_{\beta t}\right] \\
& +\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T L_{\beta t}dL_{\alpha t}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T H_{\alpha t}dL_{\beta t}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T L_{\beta t}dH_{\alpha t}\right] \\
& +\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T H_{\beta t}dL_{\alpha t}\right]+\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T L_{\alpha t}dH_{\beta t}\right].
\end{aligned}$$

Let us denote the k-th term on the right-hand side by I_k. It is observed that the following pair of terms are symmetric in α and β : $I_1\&I_2$, $I_3\&I_4$, $I_5\&I_7$, and $I_6\&I_8$. We make use of Lemma 6.11 and Lemma 6.13 repeatedly in what follows in order to transform the

terms of the partial derivatives with respect to x and y inside the expectations above, to the partial derivatives with respect to y only.

We show the derivation for I_1 and I_2 below. Noting that $dH_{\beta t} = \rho_{\beta t}\frac{\upsilon_{\beta 1,t}}{2(\upsilon_{\beta 0,t})^{1/2}}dB_t^{\beta} - \frac{1}{2}\rho_{\beta t}^2\upsilon_{\beta 1,t}dt$, we have

$$\begin{aligned} I_1 &\triangleq \mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\int_0^T H_{\alpha t}dH_{\beta t}\right] \\ &= \mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\int_0^T H_{\alpha t}\rho_{\beta t}\frac{\upsilon_{\beta 1,t}}{2(\upsilon_{\beta 0,t})^{1/2}}dB_t^{\beta}\right] - \mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\int_0^T H_{\alpha t}\rho_{\beta t}^2\frac{\upsilon_{\beta 1,t}}{2}dt\right]. \end{aligned}$$

By applying Lemma 6.11 on the first term and then Lemma 6.13, we have

$$I_1 = \mathbb{E}\left[\left(\frac{\partial^3}{\partial x^3} - \frac{\partial^2}{\partial x^2}\right)P\int_0^T H_{\alpha t}\rho_{\beta t}^2\frac{\upsilon_{\beta 1,t}}{2}dt\right] = \mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T H_{\alpha t}\rho_{\beta t}^2\upsilon_{\beta 1,t}dt\right].$$

To further simplify, the substitution of the definition of $H_{\alpha t}$ yields

$$\begin{aligned} I_1 = &\frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T \rho_{\beta t}^2\upsilon_{\beta 1,t}\left(\int_0^t \rho_{\alpha s}\frac{\upsilon_{\alpha 1,s}}{(\upsilon_{\alpha 0,s})^{1/2}}dB_s^{\alpha}\right)dt\right] \\ &-\frac{1}{2}\mathbb{E}\left[\frac{\partial^2 P}{\partial x\partial y}\int_0^T \rho_{\beta t}^2\upsilon_{\beta 1,t}\left(\int_0^t \rho_{\alpha s}^2\upsilon_{\alpha 1,s}ds\right)dt\right]. \end{aligned}$$

By applying the Fubini Theorem and Lemma 6.11, the first term is equal to

$$\frac{1}{2}\mathbb{E}\left[\frac{\partial^3 P}{\partial x^2\partial y}\int_0^T \rho_{\beta t}^2\upsilon_{\beta 1,t}\left(\int_0^t \rho_{\alpha s}^2\upsilon_{\alpha 1,s}ds\right)dt\right].$$

Finally, by Lemma 6.13 we have

$$I_1 = \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \rho_{\beta t}^2\upsilon_{\beta 1,t}\left(\int_0^t \rho_{\alpha s}^2\upsilon_{\alpha 1,s}ds\right)dt\right].$$

As I_1 and I_2 is symmetric in (α,β), we have

$$I_2 = \mathbb{E}\left[\frac{\partial^2 P}{\partial x^2}\int_0^T H_{\beta t}dH_{\alpha t}\right] = \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \rho_{\alpha t}^2\upsilon_{\alpha 1,t}\left(\int_0^t \rho_{\beta s}^2\upsilon_{\beta 1,s}ds\right)dt\right].$$

Other terms I_3 through I_8 can be obtained in a similar manner and all terms are given by

$$\begin{aligned}
I_1 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \rho_{\beta t}^2 \upsilon_{\beta 1,t}\left(\int_0^t \rho_{\alpha s}^2 \upsilon_{\alpha 1,s} ds\right) dt\right], \\
I_2 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \rho_{\alpha t}^2 \upsilon_{\alpha 1,t}\left(\int_0^t \rho_{\beta s}^2 \upsilon_{\beta 1,s} ds\right) dt\right], \\
I_3 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left(1-\rho_{\beta t}^2\right) \upsilon_{\beta 1,t}\left(\int_0^t \left(1-\rho_{\alpha s}^2\right) \upsilon_{\alpha 1,s} ds\right) dt\right], \\
I_4 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left(1-\rho_{\alpha t}^2\right) \upsilon_{\alpha 1,t}\left(\int_0^t \left(1-\rho_{\beta s}^2\right) \upsilon_{\beta 1,s} ds\right) dt\right], \\
I_5 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left(1-\rho_{\beta t}^2\right) \upsilon_{\beta 1,t}\left(\int_0^t \rho_{\alpha s}^2 \upsilon_{\alpha 1,s} ds\right) dt\right], \\
I_6 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \rho_{\alpha t}^2 \upsilon_{\alpha 1,t}\left(\int_0^t \left(1-\rho_{\beta s}^2\right) \upsilon_{\beta 1,s} ds\right) dt\right], \\
I_7 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left(1-\rho_{\alpha t}^2\right) \upsilon_{\alpha 1,t}\left(\int_0^t \rho_{\beta s}^2 \upsilon_{\beta 1,s} ds\right) dt\right], \\
I_8 &= \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \rho_{\beta t}^2 \upsilon_{\beta 1,t}\left(\int_0^t \left(1-\rho_{\alpha s}^2\right) \upsilon_{\alpha 1,s} ds\right) dt\right].
\end{aligned}$$

Summing up these terms then gives the result (6.17).

6.7.4 Proof of (6.23)

Take $i=\alpha$ and $j=\beta$ such that $\alpha \neq \beta$. We observe that the two terms in $\Phi_T^{\alpha,\beta}$ of (6.17) are symmetric in (α,β). Hence, (6.17) is equivalent to

$$\Phi_T^{\alpha,\beta} = \gamma(\alpha,\beta) + \gamma(\beta,\alpha),$$

where

$$\gamma(\alpha,\beta) = \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left[\int_0^t \upsilon_{\alpha 1,s} ds\right] \upsilon_{\beta 1,t} dt\right] = \mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T e^{-\kappa_\beta t}\left[\int_0^t \upsilon_{\alpha 1,s} ds\right] e^{\kappa_\beta t}\upsilon_{\beta 1,t} dt\right].$$

By applying Lemma 6.12 with $f(t) = e^{-\kappa_\beta t}$ and $Z(t) = \left(\int_0^t \upsilon_{\alpha 1,s} ds\right) e^{\kappa_\beta t}\upsilon_{\beta 1,t}$ we have

$$\begin{aligned}
\gamma(\alpha,\beta) &= \underbrace{\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left(\int_t^T e^{-\kappa_\beta s} ds\right) e^{\kappa_\beta t}\upsilon_{\alpha 1,t}\upsilon_{\beta 1,t} dt\right]}_{\mathcal{I}_1} \\
&\quad + \underbrace{\mathbb{E}\left[\frac{\partial^2 P}{\partial y^2}\int_0^T \left(\int_t^T e^{-\kappa_\beta s} ds\right)\left(\int_0^t \upsilon_{\alpha 1,u} du\right) e^{\kappa_\beta t}\xi_{\beta t}\sqrt{\upsilon_{\beta 0,t}} dB_t^\beta\right]}_{\mathcal{I}_2}.
\end{aligned}$$

For $\mathcal{I}_1$, we can directly apply the second equality in Lemma 6.14 by taking $h(t) = \left(\int_t^T e^{-\kappa_\beta s} ds\right) e^{\kappa_\beta t} = e^{\kappa_\beta t} \omega_{t,T}^{(-\kappa_\beta,1)}$, and readily obtain

$$\mathcal{I}_1 = \left(\omega_{0,T}^{(\kappa_\alpha,\phi_{\alpha 1}),(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\alpha,1),(-\kappa_\beta,1)} + \omega_{0,T}^{(\kappa_\beta,\phi_{\beta 1}),(\kappa_\alpha,\phi_{\alpha 1}),(-\kappa_\alpha,1),(-\kappa_\beta,1)}\right) \mathbb{E}\left[\frac{\partial^4 P}{\partial x^2 \partial y^2}\right]$$

because

$$\omega_{t,T}^{(-(\kappa_\alpha+\kappa_\beta),h)} = \int_t^T e^{-(\kappa_\alpha+\kappa_\beta)u} \left(\int_u^T e^{-\kappa_\beta s} ds\right) e^{\kappa_\beta u} du = \omega_{t,T}^{(-\kappa_\alpha,1),(-\kappa_\beta,1)}.$$

For the second term $\mathcal{I}_2$, we apply Lemma 6.11 along with the Fubini Theorem, and then by Lemma 6.12, we have

$$\begin{aligned}
\mathcal{I}_2 &= \mathbb{E}\left[\frac{\partial^3 P}{\partial x \partial y^2} \int_0^T \omega_{t,T}^{(-\kappa_\beta,1)} e^{\kappa_\beta t} \rho_{\beta t} \xi_{\beta t} \upsilon_{\beta 0,t} \left(\int_0^t \upsilon_{\alpha 1,u} du\right) dt\right] \\
&= \mathbb{E}\left[\frac{\partial^3 P}{\partial x \partial y^2} \int_0^T \left(\int_0^t e^{\kappa_\beta t} \phi_{\beta 1}(t) \omega_{t,T}^{(-\kappa_\beta,1)} dt\right) \upsilon_{\alpha 1,u} du\right] \\
&= \mathbb{E}\left[\frac{\partial^3 P}{\partial x \partial y^2} \int_0^T \omega_{t,T}^{(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\beta,1)} \upsilon_{\alpha 1,t} dt\right].
\end{aligned}$$

Then, we make use of the first equality in Lemma 6.14 by taking $h(t) = \omega_{t,T}^{(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\beta,1)}$. Since it holds that by definition

$$\omega_{t,T}^{(-\kappa_\alpha,h)} = \int_t^T e^{-\kappa_\alpha u} \omega_{u,T}^{(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\beta,1)} du = \omega_{t,T}^{(-\kappa_\alpha,1),(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\beta,1)},$$

we see

$$\mathcal{I}_2 = \omega_{0,T}^{(\kappa_\alpha,\phi_{\alpha 1}),(-\kappa_\alpha,1),(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\beta,1)} \mathbb{E}\left[\frac{\partial^4 P}{\partial x^2 \partial y^2}\right].$$

As a result, we obtain

$$\gamma(\alpha,\beta) = C(\alpha,\beta)\, \mathbb{E}\left[\frac{\partial^4 P}{\partial x^4 \partial y^2}\right],$$

where

$$\begin{aligned}
C(\alpha,\beta) &= \omega_{0,T}^{(\kappa_\alpha,\phi_{\alpha 1}),(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\alpha,1),(-\kappa_\beta,1)} + \omega_{0,T}^{(\kappa_\beta,\phi_{\beta 1}),(\kappa_\alpha,\phi_{\alpha 1}),(-\kappa_\alpha,1),(-\kappa_\beta,1)} \\
&\quad + \omega_{0,T}^{(\kappa_\alpha,\phi_{\alpha 1}),(-\kappa_\alpha,1),(\kappa_\beta,\phi_{\beta 1}),(-\kappa_\beta,1)}.
\end{aligned}$$

Combining the results, we are able to arrive at (6.23) in Theorem 6.3.

Part III

Empirical Analysis

Chapter 7

Non-linear Term Structure Modeling near Zero Lower Bound

7.1 Introduction

The Gaussian affine term structure model (ATSM) has been a popular choice in the modeling of yield curve given its analytical tractable bond pricing formula as well as the linear dependence of the model-implied bond yields to the underlying factors or state variables (Piazzesi, 2010). The model allows one to summarize the complex movements of bond yields into a small number of factors while imposing the no-arbitrage restrictions among bond yields with different maturities. Traditionally, these factors are regarded as *latent* and are usually related to the first three principal components of bond yields, including the level, slope and curvature factors. From an economic perspective, bond yields should interact closely with the macroeconomy and it is very tempting to relate these factors driving bond yields to various macroeconomic variables such as measures of inflation, real activity and monetary stance. This exercise of linking bond yields to macroeconomic variables, called the macro-finance term structure modeling, allows researchers to explain the movements in bond yields with a richer economic interpretation and potentially improve the prediction of future bond yields by incorporating information beyond the bond market. There have been a number of papers that explore the role of macroeconomic variables in the arbitrage-free term structure modeling. Ang and Piazzesi (2003) employ two measures of inflation and real activity and find that these macroeconomic factors explain up to 85% of the time-series variation of bond yields. Diebold et al. (2006) study the dynamic interaction between the macroeconomy and the yield curve. They find that macro variables strongly affect future movements in the yield curve with a feedback from the yield curve to the macroeconomy. Ang et al.

(2006) explore the Taylor rule interpretation of a macro-finance model by taking the inflation rate and output gap as the state variables. Li et al. (2013) extend the idea to model a time-varying Taylor rule by incorporating regime-dependent policy response coefficients. The monograph Diebold and Rudebusch (2013) provide a succinct summary on the recent development in term structure modeling with macro-finance features.

Despite its popularity in the macro-finance literature, there is one major shortcoming of the Gaussian ATSM: it does not constraint the interest rate to be non-negative.[1] This may be problematic for the prediction of future bond yields when interest rates are very close to the zero lower bound, such as the cases of the Japanese government bond (JGB) yields since 1995 and the US treasury yields after the financial crisis of 2008. Against this background, an alternative is the quadratic term structure model (QTSM) as advocated by Ahn et al. (2002) and Leippold and Wu (2002), which naturally accommodates non-negative interest rates. Indeed, the quadratic models have been widely adopted by market participants in the pricing and hedging of interest rate derivatives given its nice analytical tractability and the guarantee of non-negative model-implied interest rates (Andersen and Piterberg, 2010; Kijima et al., 2009b). However, there have been very few formal empirical studies on the QTSM in particular its performance under the zero interest rate policy. Until recently, using a two-factor yield-only model, Kim and Singleton (2012) and Andreasen and Meldrum (2013) demonstrate the strength of the QTSM model in the statistical description of the yield curve data for the JGB yields and the US treasury yields, respectively.

In this chapter, we study the pooling prediction of the future bond yields (term structure) of the Gaussian ATSM and QTSM with macro-finance features. The exposition is divided into two parts. Firstly, we estimate the two macro-finance term structure models using the Bayesian Markov Chain Monte Carlo (MCMC) procedure as in Ang et al. (2011). The procedure is suitable for the estimation of a non-linear state space model as in the case of QTSM due to the quadratic dependence of the bond yields to the state variables. The main objective is to compare the empirical performance, both in terms of the cross-sectional fitting of bond yields and in-sample forecasting performance, of the two macro-finance term structure models under the zero interest rate policy in the JGB market. To our knowledge, this is the first empirical study to compare the ATSM and the QTSM under the macro-finance setting and the low interest rate environment simultaneously. In the second part, we study the optimal pooling of the two models and attempt to derive a better forecast of the term structure by combining the advantages of the two models on hand. The idea of optimal prediction pool is pioneered by Geweke

[1] It is noted that another class of ATSM built on the square root process (Cox et al., 1985) is not suitable for macro-finance modelings because the state variables are positive by construction. This is in contrast to the fact that most of the macroeconomic variables can take negative values (e.g., inflation and output gap).

and Amisano (2011, 2012) in which potentially misspecified models are pooled together in order to improve the prediction accuracy using a log score criteria. In particular, the authors assume that neither of the two competing models are correctly specified. In the context of this chapter, the two term structure models are potentially misspecified due to the following structural reasons:

1. The Gaussian ATSM is naturally linked to the macroeconomic variables in a linear fashion and empirical studies have shown that it provides a good-fit to the macroeconomic variables and bond yields simultaneously in normal times. However, the model does not constraint non-negative interest rates which can be problematic when the zero lower bound is binding.

2. The QTSM naturally precludes negative interest rates. However, it is not clear whether the enforced non-linear mapping of the bond yields to the macroeconomic variables would provide a good fit to the data and being able to generate a reasonable prediction density.

We adopt three different novel approaches, with increasing complexity, in modeling the weighting coefficient that pools the bond yield prediction densities of the two models. In particular, the two later approaches equipped with time-varying weighting coefficient allows us to investigate the relative goodness in forecasting of the ATSM and QTSM during different sample periods (Negro et al., 2014; Waggoner and Zha, 2012). A related paper to ours is Eo and Kang (2014) who consider the model combination of the dynamic Nelson-Siegel model (DNSM) and ATSM using only latent factors.[2] We differ from Eo and Kang (2014) in focusing on the macro-finance interaction as well as the combination of two class of models in which the prediction densities are substantially different when interest rates are near the zero lower bound.

Our estimation results show that the QTSM dominates in its forecasting performance when interest rates are close to zero, while the ATSM provides a better fitting of the bond yields and macro factors simultaneously in other periods. It is worth to note that the ATSM predicts negative interest rate with almost 40% to 50% of the probability when the JGB yields are close to zero since late 1995. This indicates the importance to take into account the zero lower bound when interest rates are low. As both the ATSM and QTSM with macro-finance features can be potentially misspecified, it is recommended that one should use a combination of the two models in the prediction of future bond yields. Although we focus on the JGB data, the empirical results shed light on the future research on macro-finance modeling using the US data given the Federal Reserve's zero interest rate policy since December 2008.

[2]The DNSM can be described as a sub-class of the Gaussian ATSM with certain parameter restrictions.

7.2 Term Structure Models

7.2.1 Setup

We adopt a discrete time setting for the macro-finance term structure modeling. All the data are of quarterly frequency and hence we can interpret a single period to be one quarter. The key ingredient in the term structure modeling is the linkage between the short-rate r_t and the Gaussian state vector X_t taking values in $\mathbb{R}^M$ as

$$\begin{aligned} r_t &= \phi(X_t), \\ X_{t+1} &= \mu^Q + \Phi^Q X_t + \Sigma\varepsilon_{t+1}, \end{aligned}$$

with $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}_{M\times M})$, μ^Q is a $M \times 1$ vector and Φ^Q is a $M \times M$ matrix. The notation Q denotes the risk-neutral probability measure. Without much loss of generality, we can specify the market price of risk as

$$\lambda_t = \lambda_0 + \lambda_1 X_t,$$

where λ_0 is a $M \times 1$ vector and λ_1 is a $M \times M$ matrix. Hence, the real-world dynamics of the state vector is given by

$$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma\varepsilon_{t+1},$$

with

$$\mu^Q = \mu^P - \Sigma\lambda_0, \qquad \Phi^Q = \Phi^P - \Sigma\lambda_1,$$

where P denotes the real-world measure (Wright, 2011; Ang et al., 2011). The corresponding pricing kernel has the form

$$\xi_{t+1} = \exp\left(-r_t + \frac{1}{2}\lambda_t^T\lambda_t - \lambda_t^T\varepsilon_{t+1}\right)\xi_t,$$

and the time$-t$ price of a n-period zero-coupon bond can be formulated as

$$P_t^n = \mathbb{E}_t^P\left[\frac{\xi_{t+1}}{\xi_t}P_t^{n-1}\right] = \mathbb{E}_t^Q\left[\exp\left(-\sum_{i=0}^{n-1} r_{t+i}\right)\right].$$

We can also compute the n-period bond yield as

$$y_t^n = -\frac{1}{n}\log P_t^n.$$

Under the ATSM or the QTSM specification of the short rate function $r_t = \phi(X_t)$, it is possible to derive the bond pricing formula in terms of a recursive relationship. In

continuous-time modeling, this corresponds to a system of ordinary differential equation that determines the bond prices. We refer the readers to Piazzesi (2010) for continuous-time affine model and Ahn et al. (2002) for continuous-time quadratic Gaussian model.

7.2.2 Affine term structure model

The Gaussian ATSM is specified as

$$r_t = \delta_0 + \delta_1^T X_t, \tag{7.1}$$

i.e., the one-period short rate is a linear function to the selected factors. As the state variable X_t is Gaussian, there is no guarantee that the short rate to be non-negative. The bond pricing formula follows from Duffie and Kan (1996) as

$$P_t^n = \exp\left(A_n + B_n^T X_t\right), \tag{7.2}$$

where A_n is a scalar and B_n is a $M \times 1$ vector satisfying the recursive relationship

$$\begin{aligned} A_n &= -\delta_0 + A_{n-1} + B_{n-1}^T \mu^Q + \frac{1}{2} B_{n-1} \Sigma\Sigma^T B_{n-1}^T, \\ B_n^T &= -\delta_1^T + \Phi^Q B_{n-1}^T, \end{aligned} \tag{7.3}$$

for $n = 1, 2, ..., N$ with $A_1 = -\delta_0$ and $B_1 = -\delta_1$. As a result, the model-implied bond yield is a linear function to the state variable X_t as

$$y_t^n = -\frac{1}{n} \log P_t^n = a_n + b_n^T X_t, \tag{7.4}$$

by taking $a_n = -A_n/n$ and $b_n = -B_n/n$ as the factor loadings.

7.2.3 Quadratic term structure model

The QTSM is specified as

$$r_t = \alpha_0 + \beta_0^T X_t + X_t^T \Psi_0 X_t, \tag{7.5}$$

i.e., the one-period short rate is a quadratic function to the selected factors. To ensure non-negative interest rates, one need to set the initial loadings α, β and Ψ appropriately. The n-period zero coupon bond price can be formulated as

$$P_t^n = \exp\left(A_n + B_n^T X_t + X_t^T C_n X_t\right), \tag{7.6}$$

where A_n is a scalar, B_n is a $M \times 1$ vector and C_n is a $M \times M$ matrix satisfying the recursive relationship

$$\begin{aligned} A_n &= -\alpha_0 + A_{n-1} + B_{n-1}^T \mu^Q + \mu^T C_{n-1} \mu^Q - \frac{1}{2} \det\left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right) \\ &\quad + \frac{1}{2}\left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1}\mu^Q\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1} \left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1}\mu\right), \\ B_n^T &= -\beta_0^T + B_{n-1}^T \Phi^Q + 2\mu C_{n-1}\Phi^Q \\ &\quad + 2\left(\Sigma^T B_{n-1} + 2\Sigma^T C_{n-1}\mu\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1} \Sigma^T C_{n-1}\Phi^Q, \\ C_n &= -\Psi_0 + \left(\Phi^Q\right)^T C_{n-1}\Phi^Q + 2\left(\Sigma^T C_{n-1}\Phi^Q\right)^T \left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1} \left(\Sigma^T C_{n-1}\Phi^Q\right) \end{aligned} \tag{7.7}$$

for $n = 1, 2, ..., N$ with $A_1 = -\alpha_0$, $B_1 = -\beta_0$ and $C_1 = -\Psi_0$. As a result, the model-implied bond yield can be expressed as

$$y_t^n = -\frac{1}{n} \log P_t^n = a_n + b_n^T X_t + X_t^T c_n X_t \tag{7.8}$$

by taking $a_n = -A_n/n$, $b_n = -B_n/n$ and $c_n = -C_n/n$ as the factor loadings. Note that even when we set the initial loadings as $\alpha_0 = 0$ and $\beta_0 = 0_M$, the bond yield has a constant term loading and a linear term loading as a_n and b_n respectively.

7.2.4 Estimation method

Given the bond pricing formula that relates the model-implied bond yields to the selected state variables, we can formulate our estimation procedure in terms of a non-linear state-space model as follows:

- **Measurement Equation**

  The measurement equation describes the evolution of the observed bond yields $\hat{y}_t^n$ as

  $$\hat{y}_t^n = a_n + b_n^T X_t + X_t^T c_n X_t + \omega_{n,t}, \tag{7.9}$$

  with $n = 1, 2, ..., N$ and $\omega_{n,t}$ are the measurement errors which are i.i.d. normals. Moreover, we assume that the selected state variables are observed with measurement errors $\omega_{X,t}$ as

  $$\hat{X}_t = X_t + \omega_{X,t}, \tag{7.10}$$

  where $\hat{X}_t$ is the observed state variables and $\omega_{X,t}$ are i.i.d. normals.

- **State Equation**

The state equation is given by the evolution of the latent state vector X_t under the real-world measure P as

$$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma\varepsilon_{t+1}, \tag{7.11}$$

which is a standard VAR(1) system.

Therefore, (7.9), (7.10) and (7.11) together form a non-linearity state space model with $N + M$ observables (N observed bond yields and M observed state variables) and M latent factors. In the Appendix, we present the Bayesian MCMC procedure to estimate the model parameters. In the estimation, it is important to calibrate the size of the measurement errors for state variables and bond yields. After a number of trial runs, we set the measurement errors to be 2.5 bps for our quarterly data which can be translated to 10 bps for annualized data.

7.2.5 Data and factors

We use the bond yield data in Wright (2011) for the JGB market during the sample period from 1990Q1 to 2008Q3. The JGB yield curve data is obtained from Datastream and the author's calculation based on the Svensson interpolation methodology. In order to keep the consistency with previous macro-finance studies, we use the JGB yields of the 1, 4, 8, 12, 16 and 20 quarters. We construct the term structure model with one yield-curve factor and two macroeconomic factors, including

1. The level factor (L_t): this is proxied by the short-term interest rate using the Bank of Japan's collateralized overnight call rate.
2. Measure of real activity (g_t): based on the exponentially weighted moving average of quarterly GDP growth.
3. Measure of inflation (π_t): based on the exponentially weighted moving average of quarterly inflation.

The overnight call rate is obtained from the Bank of Japan's website, while the last two macroeconomic variables are obtained from the dataset of Wright (2011). The two selected marco variables $X_t = (y_t, \pi_t)$ are widely taken to be a set of fundamentals that capture the macroeconomic dynamics. A similar set of marcoeconomic variables have been employed in Ang and Piazzesi (2003) and Bernanke et al. (2004).

Recent empirical studies find that it is important to incorporate the yield curve factor in order to capture both the time-series and cross-sectional variations of the term structure

dynamics (Diebold et al., 2006; Joslin et al., 2013, 2014). Indeed, the seminal macro-finance model that imposes the Taylor-rule mechanism in which the short-rate is a linear function of output and inflation, is found to be insufficient in capturing simultaneously the time-series and cross-sectional variations of the bond yields, unless a time-varying Taylor rule is allowed (Ang et al., 2011; Li et al., 2013). In our estimation, we find that the incorporation of the yield-curve level factor significantly improve the bond yield fitting quality while allowing us to maintain the model dimensionality to be less than three. Moreover, as the short-term interest rate is taken as a proxy to the level factor, it also bears the macro-economic information on the central bank's monetary stance as noted in Bernanke et al. (2004).

We assume that the instantaneous short-rate depends only on the yield-curve level factor but not the macro-economic factors, that is $r_t = \phi(L_t)$. There are several practical reasons for this specification. Firstly, in the case of ATSM we can simply take $\delta_0 = 0$ and $\delta_1 = (1, 0, 0)$. This avoids the need to estimate the loading coefficients of the short-rate function which is prone to identification problem when measurement errors are included. This is also a common setup adopted in previous studies such as Bernanke et al. (2004) and Ang et al. (2011). Secondly, in the case of QTSM, this allows an easier parameterization of the short-rate function in order to avoid non-negative interest rate, because imposing a second-order polynomial to be non-negative can be done analytically and is easy-to-visualize.[3] Thirdly, this allows a fair comparison in between the ATSM and QTSM because the short-rate function under QTSM has a larger number of free parameters (13 loading coefficients) than the ATSM (4 loading coefficients). Furthermore, our estimated results confirms that the specifications with parameter restrictions on the loadings are flexible enough to provide a high quality on the fitting of bond yields.

We make a remark that even though there is no interaction in between the instantaneous short-rate r_t and the macro-economic factors g_t and π_t, the macro-economic factors can still influence the medium to longer term bond yields through the VAR-dynamics under the risk-neutral measure Q. To be specific, recall that the model-implied bond yields under ATSM and QTSM are given by $y_t^n = a_n + b_n^T X_t$ and $y_t^n = a_n + b_n^T X_t + X_t^T c_n X_t$ respectively. As long as the relevant entries in a_n, b_n and c_n are non-zero, the model-implied bond yield are to be affected by the macro-economic factors under the risk-neutral measure Q.

To facilitate the understanding of the estimation results, let us review briefly the Bank of Japan's monetary policy since the 1990s. The Bank of Japan started to ease the base interest rate (the uncollateralized overnight call rate) in the early 1990s, which

[3]It becomes more difficult to visualize the short-rate function as the dimensionality increases. For example, to ensure the non-negativity condition when the short-rate is loaded to two factors, one needs to calibrate a quadratic surface over the zero plane.

is subsequently lowered down to 0.5 percent and 0.25 percent in 1995Q4 and 1998Q4 respectively. To further stimulate the economy, the Bank of Japan adopted the zero interest rate policy (ZIRP) during the period from 1999Q1 to 2000Q3 by keeping the base interest rate effectively to zero. After a short-term recovery in early 2000s, the Japan economy returned to a recession against the backdrop of the internet bubble, which led to the introduction of the quantitative monetary easing policy (QMEP) in order to combat the deflationary pressure. Since then, the Japanese base interest rate has been kept very close to the zero lower bound. Baba (2006) provides a comprehensive review of the Bank of Japan monetary policy and the JGB market development over the sample period.

7.3 Estimation Results: Term Structure Models

7.3.1 Model estimation

First-of-all, we look at the filtered factors and model-implied bond yields to investigate the goodness-of-fit of the Gaussian ATSM and QTSM. Table 7.1 reports the summary statistics of the posterior estimations of the ATSM parameters, while Figure 7.1 shows the posterior distributions of the filtered factors and the model-implied bond yields. These estimations are obtained by 10,000 draws of MCMC samplings after discarding the first 5000 burn-in draws based on the Bayesian methods described in the Appendix. The solid blue line and the dashed red represent actual values and fitted values, respectively, the dashed blue line represents discrepancy between them, and the shaded grey band represents 90% confidence interval of the distribution. It can be seen that macro factors track quite closely to the actual data and the fitting of the bond yields are reasonably good across maturities. Apparently, the ATSM is adequate to model the joint dynamics in macro factors and bond yields. Nevertheless, it is noted that the posterior distributions of the model-implied bond yields often bleach the zero lower bound with a noticeable probability of negative interest rates during the sample periods after late 1995. This generates a large degree of pricing errors when the actual short-term bond yields are effectively zero under the ZIRP.

Table 7.2 and Figure 7.2 report the corresponding results for the QTSM which imposes non-negative interest rates and introduces a non-linear mapping in between the factors and the bond yields. In Figure 7.2, we see that the filtered output and inflation factors also track closely to the actual data. Furthermore, as the model-implied bond yields of QTSM are guaranteed to be positive, it is able to provide a better fit to the very short-term bond yields near the zero lower bound (1Q and 4Q). However, the fitting to

long-term bond yields (16Q and 20Q) is less satisfactory in the case of QTSM, which suggests that the quadratic mapping in between the factors and the bond yields maybe prone to misspecification.

To better understand how the term structure model predicts the responses of bond yields to shocks in the underlying macro variables (i.e., impulse response), it is important to take a closer look at the estimated factor loadings. Figure 7.3 reports the factor loadings of the estimated Gaussian ATSM using the recursive relationship (7.3) and the posterior means under Q-measure as reported in Table 7.1. Recall that the short-term interest rate is taken as one of the state variables and we have imposed the initial loading of the inflation and output factors: $\delta_{1,2}$ and $\delta_{1,3}$ to be zero as in Ang et al. (2011). For the ATSM, we see that the output and inflation loadings: b_2 and b_3, are positive for all maturities, which is consistent with the economic implications of the Taylor rule. For example, a positive shock to the output gap induces an upward shift and a steepening of the yield curve, which reflects that the slope of yield curve is highly related to economic outlook (Diebold and Rudebusch, 2013). As expected, the loading to the yield-curve level factor: b_1, is less than one. This indicates that the transmission effect of the short-term interest rate (monetary stance) to the long-end of the yield curve is imperfect as can be seen from Figure 7.3.

Figure 7.4 reports the factor loadings of the estimated QTSM using the recursive relationship (7.7) and the posterior means under Q-measure in Table 7.2. In contrast to the ATSM which only has 4 factor loadings, the QTSM provides in total 10 loading combinations to the three factors, including 4 loadings through the linear terms: a, b_1, b_2, b_3, and 6 loadings through the quadratic terms : c_{11}, c_{22}, c_{33}, c_{12}, c_{13}, c_{23}.[4] As discussed, we set the initial loadings to the output and inflation factors: Φ_{22} and Φ_{33}, and other off-diagonal elements in (7.5), to zero as reflected in Figure 7.4(b). Firstly, it is worth to take a look at the diagonal elements of the factor loading c_n, which captures most of the variation in the yield curve. As expected, the loading to the quadratic terms of the inflation and output factors: c_{22} and c_{33}, are positive, indicating that investors demand a higher bond yields when inflation uncertainty and output uncertainty (second moments) are high as shown in Figure 7.4(b). In addition, the QTSM allows a flexible interaction in between different factors through the cross terms (i.e., the off-diagonal elements in the factor loading c_n). Diebold and Rudebusch (2013) note that the negative interaction in between factors can play an important role to model interest rates near the zero lower bound. In our case, the factor loadings for the cross terms of (f_t, π_t) and (g_t, π_t) : c_{13} and c_{23}, are estimated to be negative, which indicate the enhanced flexibility of the QTSM in relating bond yields to the selected macro factors. We can argue that it

[4]Note that the factor loading c_n is symmetric by construction. Accordingly, $c_{12} = c_{21}$, $c_{13} = c_{31}$, and $c_{23} = c_{32}$.

TABLE 7.1: Posterior estimates of the model parameters for ATSM. The reported values for the parameters μ and $(\Sigma\Sigma^T)_{ij}$ are multiplied by 10,000.

	Mean	90 Percentile	10-Percentile	Std. Dev.
	VAR(1)-system under P-measure			
Φ_{11}	0.8672	0.9196	0.8128	0.0428
Φ_{12}	0.0438	0.1318	-0.0457	0.0696
Φ_{13}	0.1014	0.2287	-0.0246	0.1000
Φ_{21}	0.1210	0.2150	0.0326	0.0716
Φ_{22}	0.5443	0.7098	0.3652	0.1333
Φ_{23}	-0.1947	0.0048	-0.3901	0.1593
Φ_{31}	0.0397	0.1141	-0.0312	0.0572
Φ_{32}	0.0565	0.2055	-0.1206	0.1271
Φ_{33}	0.5287	0.6922	0.3699	0.1257
μ_1	-0.6343	2.6977	-3.9661	2.5966
μ_2	12.1554	19.1896	5.6553	5.2071
μ_3	1.0896	8.0413	-4.4578	4.9450
	VAR(1)-system under Q-measure			
Φ_{11}	0.9383	0.9674	0.9080	0.0230
Φ_{12}	0.0691	0.0833	0.0538	0.0107
Φ_{13}	0.0138	0.0182	0.0093	0.0033
Φ_{21}	0.1336	0.1685	0.1105	0.0213
Φ_{22}	0.8502	0.9220	0.7832	0.0521
Φ_{23}	-0.0329	0.1774	-0.2265	0.1671
Φ_{31}	0.1347	0.1586	0.1132	0.0173
Φ_{32}	0.0825	0.1017	0.0642	0.0146
Φ_{33}	0.4517	0.5255	0.3747	0.0570
μ_1	-0.3371	-0.0775	-0.5338	0.1782
μ_2	8.7984	9.8244	7.6819	0.8208
μ_3	-2.2497	-1.7340	-2.9599	0.4368
	Variance Matrix			
$(\Sigma\Sigma^T)_{11}$	0.0155	0.0201	0.0114	0.0034
$(\Sigma\Sigma^T)_{12}$	-0.0051	-0.0001	-0.0106	0.0041
$(\Sigma\Sigma^T)_{13}$	-0.0018	0.0031	-0.0073	0.0041
$(\Sigma\Sigma^T)_{21}$	-0.0051	-0.0001	-0.0106	0.0041
$(\Sigma\Sigma^T)_{22}$	0.0370	0.0476	0.0272	0.0082
$(\Sigma\Sigma^T)_{23}$	-0.0001	0.0073	-0.0078	0.0060
$(\Sigma\Sigma^T)_{31}$	-0.0018	0.0031	-0.0073	0.0041
$(\Sigma\Sigma^T)_{32}$	-0.0001	0.0073	-0.0078	0.0060
$(\Sigma\Sigma^T)_{33}$	0.0296	0.0442	0.0145	0.0116

Notes:

1. The first 5,000 draws of MCMC sampling are discarded (burn-in) to guarantee convergence and then the next 10,000 draws are used for calculating the posterior means, the standard deviations (Std. Dev.), as well as the 10 and 90 percentiles.
2. The posterior mean is computed by averaging the MCMC draws.
3. Std. Dev. is computed as the sample standard deviation of the MCMC draws.

TABLE 7.2: Posterior estimates of the model parameters for QTSM. The reported values for the parameters μ and $(\Sigma\Sigma^T)_{ij}$ are multiplied by 10,000.

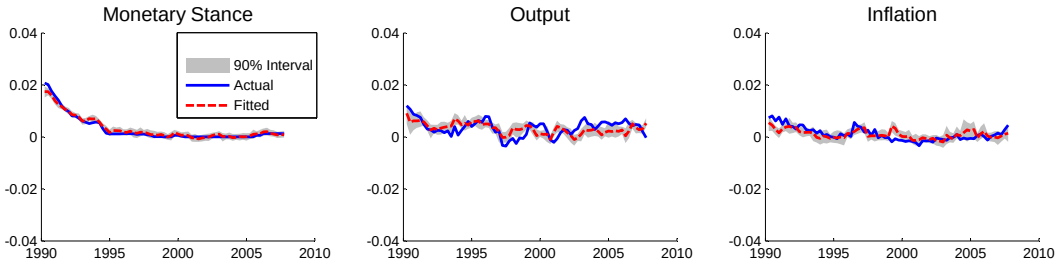
	Mean	90 Percentile	10-Percentile	Std. Dev.
	VAR(1)-system under P-measure			
Φ_{11}	0.8131	0.8934	0.7320	0.0640
Φ_{12}	0.0568	0.1696	-0.0544	0.0872
Φ_{13}	0.2270	0.3758	0.0794	0.1173
Φ_{21}	0.0169	0.1213	-0.0911	0.0827
Φ_{22}	0.6470	0.7770	0.5121	0.1035
Φ_{23}	0.0068	0.2000	-0.1789	0.1462
Φ_{31}	0.0670	0.1666	-0.0133	0.0719
Φ_{32}	0.0590	0.1544	-0.0387	0.0774
Φ_{33}	0.7496	0.9090	0.5348	0.1463
μ_1	2.2411	6.3008	-1.8670	3.2158
μ_2	7.3223	12.0844	2.6005	3.7285
μ_3	-2.4398	0.8460	-6.0533	2.8044
	VAR(1)-system under Q-measure			
Φ_{11}	0.9942	0.9993	0.9892	0.0039
Φ_{12}	0.0392	0.0484	0.0333	0.0057
Φ_{13}	0.0187	0.0215	0.0162	0.0021
Φ_{21}	0.0951	0.1243	0.0718	0.0203
Φ_{22}	0.7553	0.8453	0.6581	0.0732
Φ_{23}	-0.2717	-0.2504	-0.2947	0.0165
Φ_{31}	0.1105	0.1302	0.0922	0.0132
Φ_{32}	0.0671	0.0952	0.0479	0.0166
Φ_{33}	0.4003	0.4472	0.3493	0.0372
μ_1	-1.2593	-1.0521	-1.3488	0.1056
μ_2	7.4830	7.9915	6.9087	0.4519
μ_3	-2.4335	-2.1773	-2.7327	0.2461
	Variance Matrix			
$(\Sigma\Sigma^T)_{11}$	0.0310	0.0407	0.0225	0.0072
$(\Sigma\Sigma^T)_{12}$	0.0027	0.0110	-0.0055	0.0065
$(\Sigma\Sigma^T)_{13}$	-0.0043	0.0011	-0.0101	0.0044
$(\Sigma\Sigma^T)_{21}$	0.0027	0.0110	-0.0055	0.0065
$(\Sigma\Sigma^T)_{22}$	0.0393	0.0553	0.0268	0.0116
$(\Sigma\Sigma^T)_{23}$	-0.0070	-0.0002	-0.0141	0.0054
$(\Sigma\Sigma^T)_{31}$	-0.0043	0.0011	-0.0101	0.0044
$(\Sigma\Sigma^T)_{32}$	-0.0070	-0.0002	-0.0141	0.0054
$(\Sigma\Sigma^T)_{33}$	0.0201	0.0349	0.0097	0.0097

Notes:

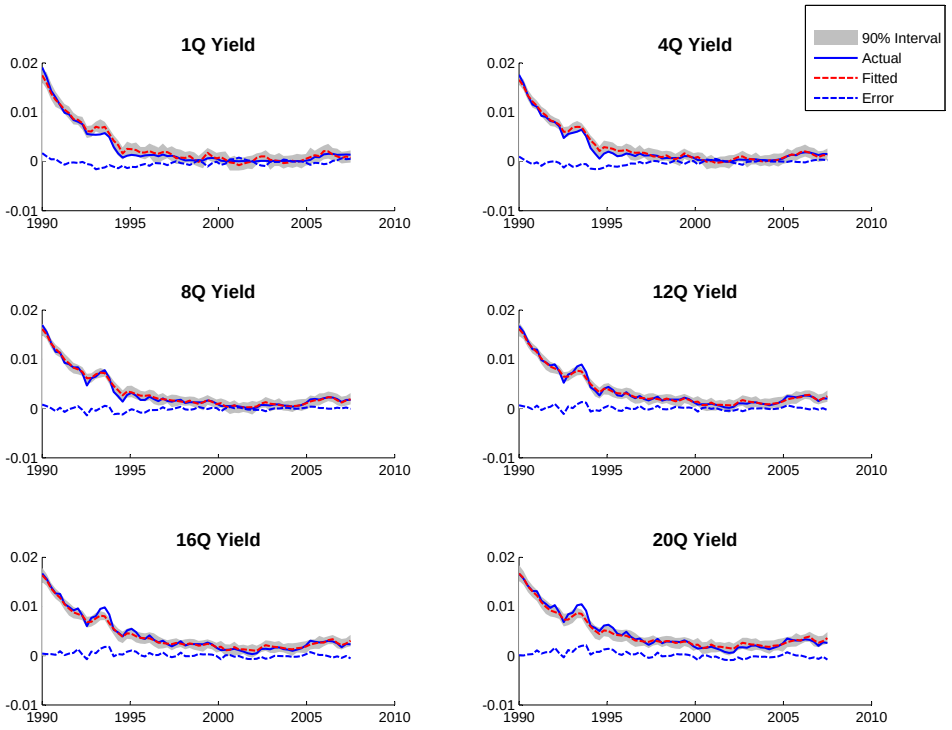
1. The first 5,000 draws of MCMC sampling are discarded (burn-in) to guarantee convergence and then the next 10,000 draws are used for calculating the posterior means, the standard deviations (Std. Dev.), as well as the 10 and 90 percentiles.
2. The posterior mean is computed by averaging the MCMC draws.
3. Std. Dev. is computed as the sample standard deviation of the MCMC draws.

FIGURE 7.1: Filtered macro factors and fitted bond yields by the ATSM

(A) Macro factors (in quarter)



(B) Bond yields (in quarter)

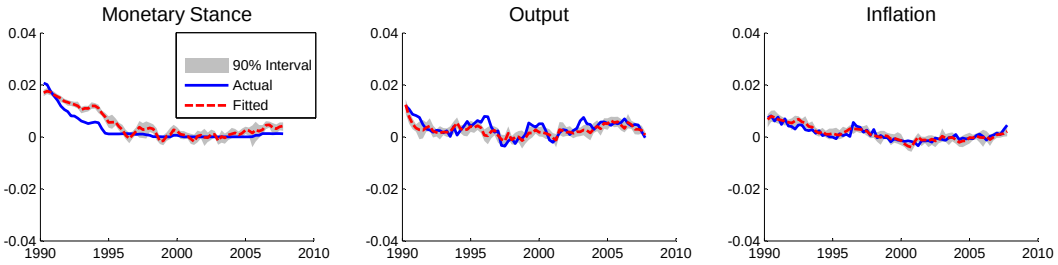


Notes:

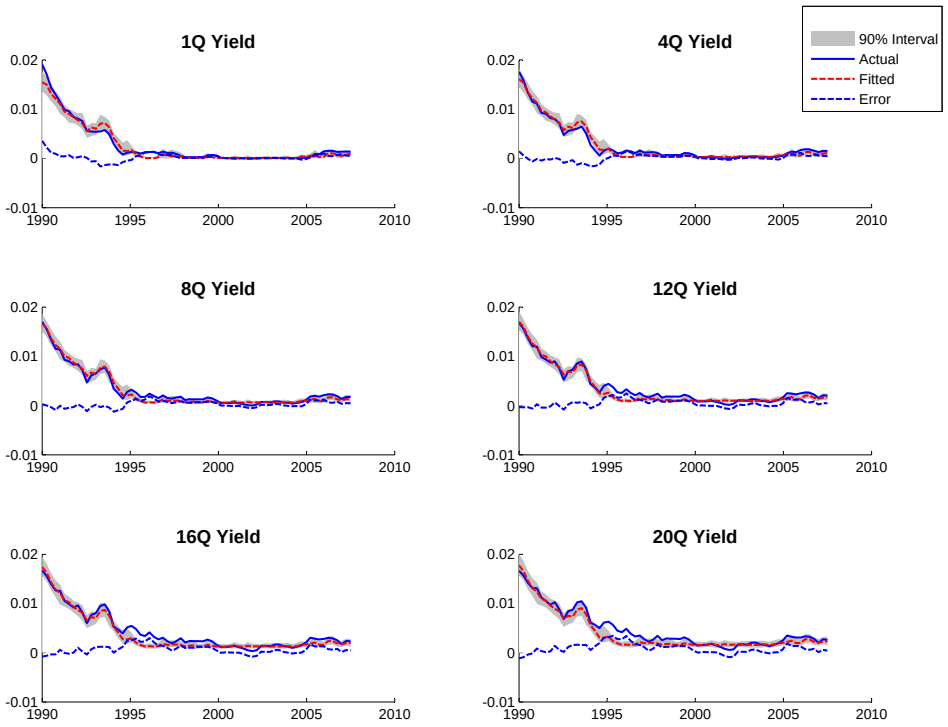
1. The solid blue line and the dashed red represent actual values and fitted values respectively, the dashed blue line represents discrepancy between them, and the shaded grey band represents 90% confidence interval of the distribution.
2. These estimations are obtained by 10,000 draws of MCMC samplings after discarding the first 5000 burn-in draws based on the Bayesian estimation described in the Appendix.

FIGURE 7.2: Filtered macro factors and fitted bond yields by the QTSM

(A) Macro factors (in quarter)



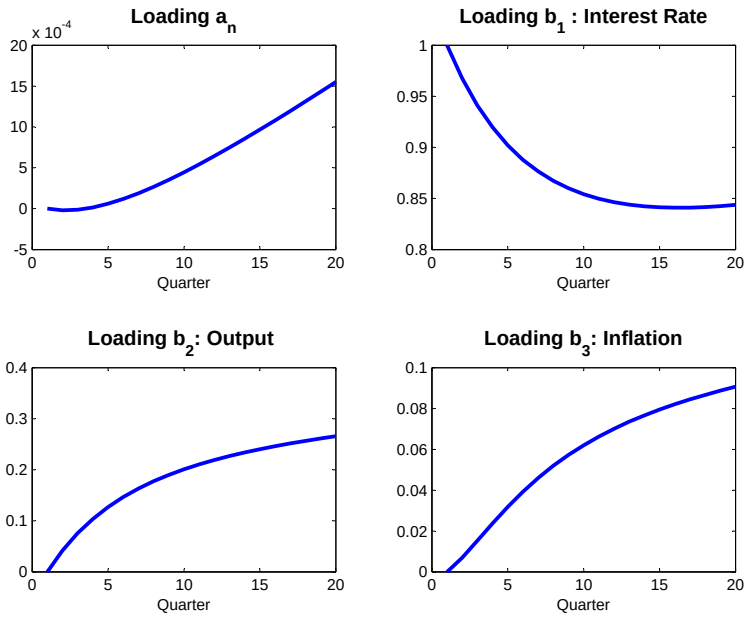
(B) Bond yields (in quarter)



Notes:

1. The solid blue line and the dashed red represent actual values and fitted values respectively, the dashed blue line represents discrepancy between them, and the shaded grey band represents 90% confidence interval of the distribution.
2. These estimations are obtained by 10,000 draws of MCMC samplings after discarding the first 5000 burn-in draws based on the Bayesian estimation described in the Appendix.

FIGURE 7.3: Factor loadings a_n and b_n for ATSM



Notes: The factor loadings of the estimated Gaussian ATSM are calculated using the recursive relationship and the posterior means under Q-measure in Table 7.1.

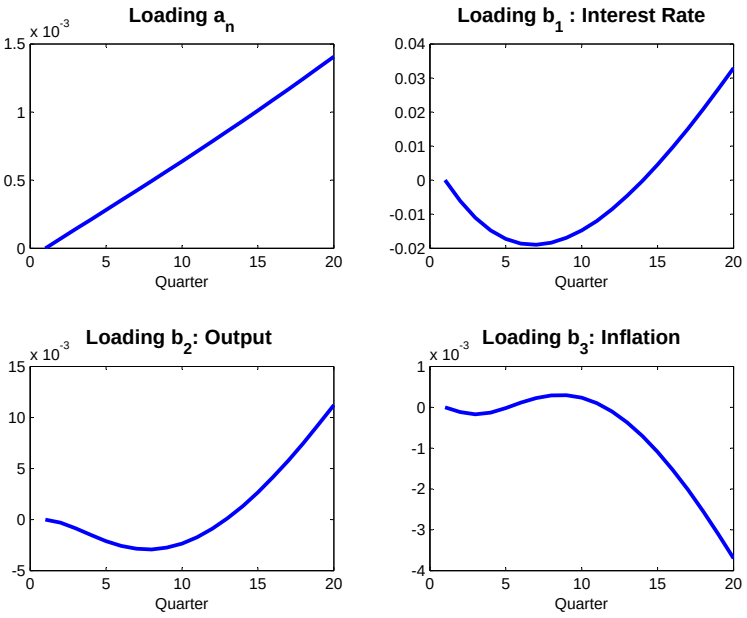
is the negative loadings of the cross terms that generate the off-setting effects such that the QTSM is able to capture the persistent and sticky short-term bond yields under the ZIRP, e.g., see the fitting of the 1Q yield in Figure 7.2(b).

7.3.2 Prediction of macro factors and bond yields

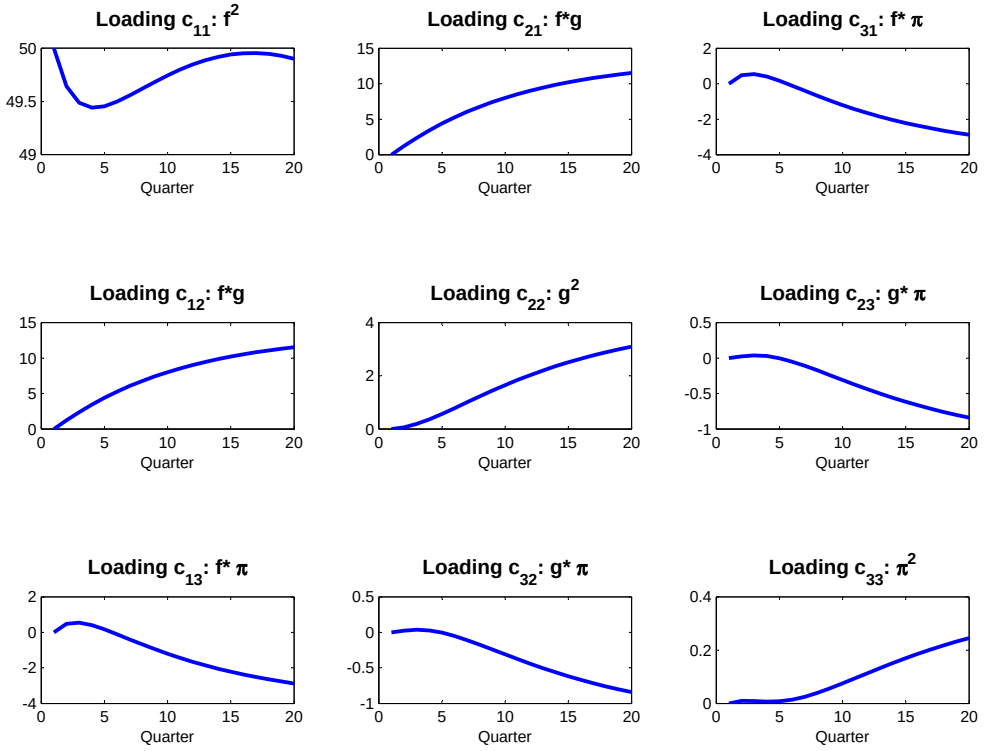
In this subsection, we study the in-sample forecasting performance of the macro factors and bond yields to analyse the models. For a more consistent explanation of the results, let us view the yield-curve level factor (short-term interest rate) as the monetary stance factor similar to the interpretation in Bernanke et al. (2004) and Diebold et al. (2006). We calculate the posterior prediction distributions of the macro factors and the JGB yields of the two models using 10,000 draws of posterior estimates over the full sample as shown in Tables 7.1 and 7.2. Figure 7.5 shows the ATSM prediction of the macro factors and the JGB yield curve across 6 maturities for the following two forecasting periods: (i) 1992Q4 - 1998Q1 and (ii) 2003Q4 - 2008Q3. The solid black line represents actual values, the solid red line represents the median of posterior prediction distributions and the shaded blue band represents 90% confidence interval of the distribution. In the period of 1992Q4 - 1998Q1, in which monetary policy has not stand under ZIRP yet, the bond yields are quite far away from the zero lower bound as in Figure 7.5(a).

FIGURE 7.4: Factor loadings a_n, b_n and c_n for QTSM

(A) Factor loadings a_n and b_n



(B) Factor loadings c_n.



Note: The factor loadings of the estimated QTSM are calculated using the recursive relationship and the posterior means under Q-measure in Table 7.2.

Although the median forecast fits well to the actual data, the ATSM predicts negative bond yields when the forecasting horizon is beyond 4 to 8 quarters. When the Bank of Japan adopted the ZIRP and the QMEP in 2003Q4, the prediction of bond yields by the ATSM is even more unrealistic: as the short-term bond yields are close to the zero lower bound, the model predicts with almost half of the probability that the bond yields are negative as in Figure 7.5(b). Even for the 5-year bond yield, there is a substantial probability of bleaching the zero lower bound when the forecasting horizon is beyond 4 quarters. This reflects that the Gaussian ATSM is very unreliable for the prediction of bond yields when interest rates are close to zero.

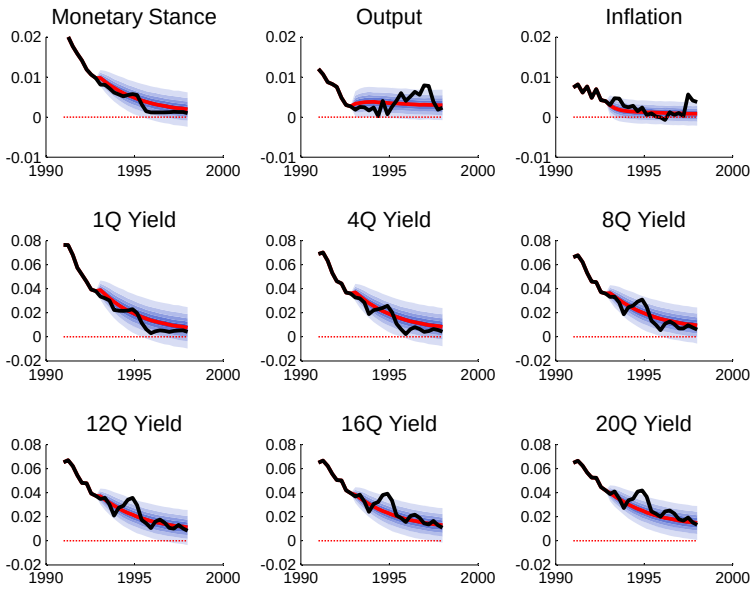
Figure 7.6 shows the QTSM prediction of the macro factors and the bond yields during the corresponding two forecasting periods. For the in-sample prediction in 1992Q4 - 1998Q1, the QTSM produces a less accurate forecast as with the ATSM model as in Figure 7.6(a). Moreover, the prediction density is positively skewed because the bond yields are bounded below by zero due to the imposition of non-negative short rate in QTSM. The strength of the QTSM is found to be prominent during the period of 2003Q4 - 2008Q3 when the zero lower bound is binding: the prediction produces only positive bond yields even though the short-term interest rate is extremely close to zero as in Figure 7.6(b). From the fan chart of the QTSM predictive density, we can observe that the probability mass near zero is significant even for medium-term to long-term forecasting horizons. This reflects the stickiness nature of the QTSM which allows one to capture the prolonged zero interest rate policy in Japan.

7.3.3 Robustness check

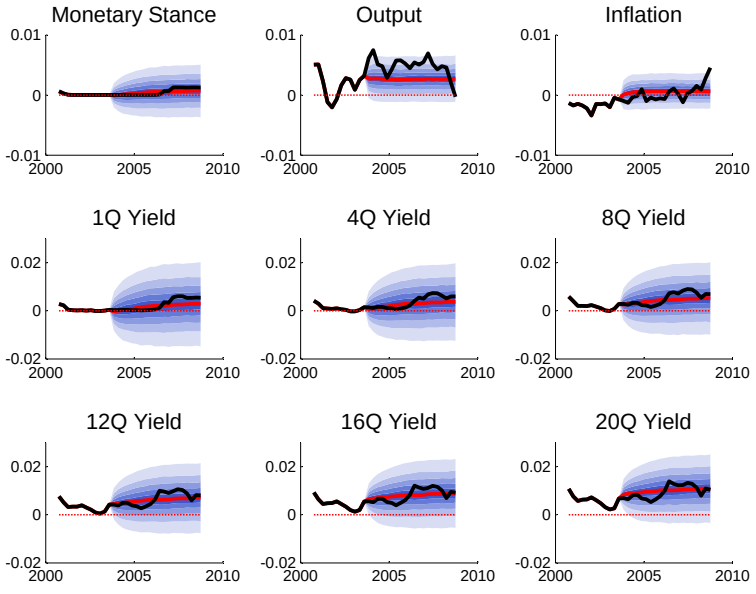
We perform two robustness checks on the sample period used for the estimation: i.) we extend the sample period before the burst of the Japan economy bubble in early 1990s by including the sample data from 1985Q1; ii.) we extend the sample period after the Lehman financial crisis by including the data until 2009Q1. In both cases, we find that the simultaneous fitting of the macro factors and the bond yields deteriorate in certain degrees. This can be explained by the structural breaks in the data set in which the time-series dynamics of the macro factors and the term structure are different under these extended sample periods. It is therefore advised to pursue a regime-switching term structure model in order to capture the structural changes in the data. Nevertheless, adding a regime-switching feature leads to a more sophisticated bond pricing formula and one has to resort to analytical approximation. To this end, we refer the readers to Li et al. (2013) for a recent empirical study on regime-switching term structure model.

FIGURE 7.5: Forecasting of macro factors and bond yields (annualized) by the ATSM

(A) Periods: 1992Q4 - 1998Q1 (under non-ZIRP).



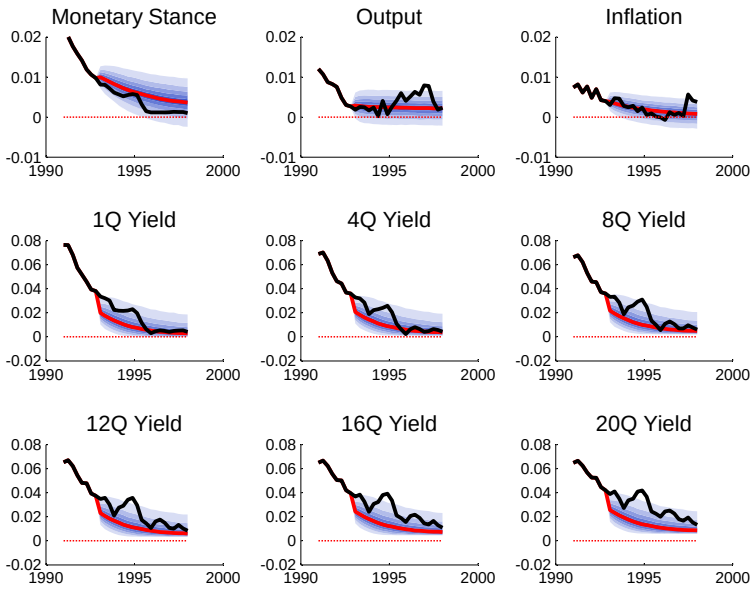
(B) Periods: 2003Q4 - 2008Q3 (under ZIRP).



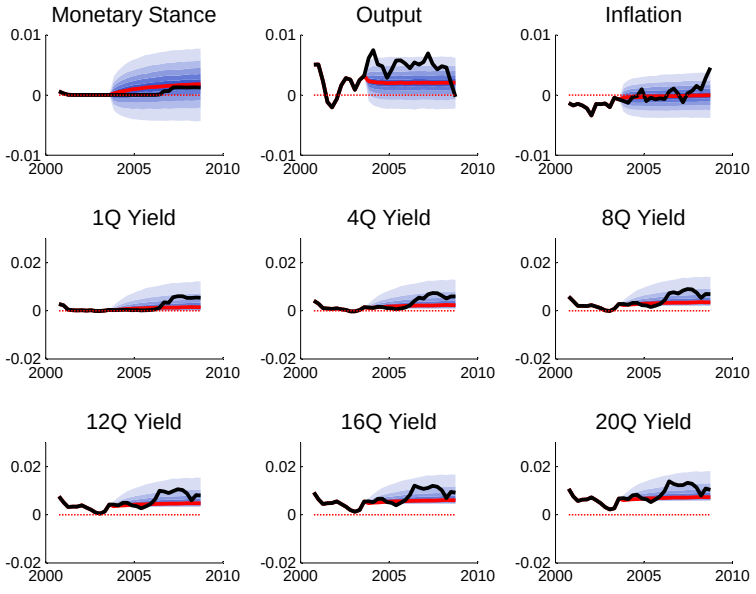
Note: The posterior prediction distributions of the macro factor and the JGB yields of the ATSM models are calculated using 10,000 draws of posterior estimates over the full sample as shown in Table 7.1.

FIGURE 7.6: Forecasting of macro factors and bond yields (annualized) by the QTSM

(A) Periods: 1992Q4 - 1998Q1 (under non-ZIRP).



(B) Periods: 2003Q4 - 2008Q3 (under ZIRP).



Note: The posterior prediction distributions of the macro factor and the JGB yields of the QTSM models are calculated using 10,000 draws of posterior estimates over the full sample as shown in Table 7.2.

7.4 Methods of Prediction Pooling

In this section, we discuss the idea of optimal prediction pooling as advocated in Geweke and Amisano (2011) and present the pooling schemes adopted in this study.

7.4.1 Motivation

From a Bayesian perspective, the *marginal likelihood* is commonly used as a criteria of model choice because it is interpreted as the prediction distribution after integrating the prior density with respect to the model parameters. Let us denote y_t a vector time series, and its history as $Y_{t-1}^O = \{y_h, ..., y_{t-1}\}$ where $h \leq 1$ is a starting date and the superscript "O" denotes the observed data. The marginal likelihood (ML) is given by

$$p^{ML}\left(y_t|Y_{t-1}^O, \mathcal{M}\right) = \int p\left(y_t|Y_{t-1}^O, \Theta, \mathcal{M}\right) p(\Theta|\mathcal{M}) d\Theta, \tag{7.12}$$

where $\mathcal{M}$ is a prediction model, $p\left(y_t|Y_{t-1}, \Theta, \mathcal{M}\right)$ and $p(\Theta|\mathcal{M})$ denote the likelihood function and the prior density of parameters Θ on a specified model $\mathcal{M}$, respectively. If we regard the marginal likelihood as a *prior* prediction distribution as noted in Geweke (2010), we can generalize and propose a *posterior* prediction distribution by replacing the prior density $p(\Theta|\mathcal{M})$ using the posterior density $p(\Theta|Y_{t-1}, \mathcal{M})$ as

$$p^{Post}\left(y_t|Y_{t-1}^O, \mathcal{M}\right) = \int p\left(y_t|Y_{t-1}^O, \Theta, \mathcal{M}\right) p(\Theta|Y_{t-1}^O, \mathcal{M}) d\Theta,$$

where $p(\Theta|Y_{t-1}, \mathcal{M})$ is the posterior density of parameters conditional on a specified model $\mathcal{M}$. Following Geweke and Amisano (2011), we use the posterior prediction distribution to construct a *log prediction scoring rule* in order to evaluate the forecasting performances using individual models as well as the calculation of the optimal weights when two models are combined for forecasting. We set up the prediction score of observation y_t^O at period t by using the posterior prediction density as

$$\underbrace{p(y_t^O;\, Y_t^O,\, \mathcal{M})}_{\text{prediction score}} = p^{Post}\left(y_t^O|Y_{t-1}^O, \mathcal{M}\right),$$

and regard it as the key element of the following all prediction pooling methods. The aim of this study is to examine various model choices and combinations in terms of the macro-finance term structure modeling using this new model selection criteria. We conduct three recent approaches of prediction pool which we explain as follows.

7.4.2 Static prediction pooling

Firstly, we discuss the static pooling based on a constant weighting as in Geweke and Amisano (2011). Given two prediction models $\mathcal{M}_1$ and $\mathcal{M}_2$, the pool of the prediction density can be constructed as the convex combination

$$p\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right) = \lambda p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + (1-\lambda)\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right), \qquad 0 \leq \lambda \leq 1,$$

with $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$ is the collection of models and $p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_i\right) = p\left(y_t^O; Y_{t-1}^O, \hat{\Theta}_i\right)$ is the prediction density with the posterior estimates $\hat{\Theta}_i$ of the model $\mathcal{M}_i$. The optimal prediction pooling is then obtained by maximizing the cumulative log prediction score LPS^{SP} as

$$LPS^{SP}(\lambda) \equiv \sum_{t=1}^{T} \log\left[\lambda p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + (1-\lambda)\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right)\right] \tag{7.13}$$

by choosing $\lambda^* = \arg\max LPS^{SP}(\lambda)$. An important assumption as noted in Geweke and Amisano (2011) is that the two candidate prediction models have to be substantially different in terms of the functional form of their predictive densities (i.e., non-nested models). In our case, the Gaussian ATSM generates a prediction density of bond yield close to a normal distribution, while the QTSM generates an asymmetric prediction density which is bounded explicitly from zero (Kim and Singleton, 2012). Under the Bayesian framework, we can estimate the pooling scheme based on the random walk Metropolis-Hastings (MH) algorithm and sample the posterior distribution of the constant weighting λ.

7.4.3 Markov-switching prediction pooling

Waggoner and Zha (2012) extend the static prediction pool by allowing the weighting coefficient λ_t to be dependent on a regime variable s_t following a Markov chain as

$$\lambda_t = \lambda(s_t) = \begin{cases} \lambda_1, & s_t = 1 \\ \lambda_2, & s_t = 2 \end{cases},$$

in which the transition matrix Q for s_t is given by

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix},$$

where $q_{ij} = \Pr(s_t = j | s_{t-1} = i)$ with $q_{11} + q_{12} = 1$ and $q_{21} + q_{22} = 1$. Conditional on the state s_t, the pool of the prediction density can be expressed as

$$p\left(y_t^O;\, Y_{t-1}^O, \mathcal{M},\, s_t\right) = \lambda(s_t)\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + (1 - \lambda(s_t))\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right),$$

with $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$. Hence, integrating out the unobservable regime s_t, we can formulate the pooled prediction density as

$$p^{MS}\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right) = \sum_{s_t=1}^{2} p\left(y_t^O; Y_{t-1}^O, \mathcal{M},\, s_t\right) p(s_t | s_{t-1})\, p\left(s_{t-1} | Y_{t-1}^O, \mathcal{M}\right),$$

where $p(s_t | s_{t-1})$ is the transition probability of the Markov chain. To estimate the Markov-switching weighting, we follow the single-move algorithm as proposed in Albert and Chib (1993) which is discussed in the Appendix. The log prediction score of Markov-switching prediction pooling with two regimes is

$$LPS^{MS}(\lambda_1, \lambda_2) \equiv \sum_{t=1}^{T} \log\, p^{MS}\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right). \tag{7.14}$$

An advantage of using the Markov-switching modeling for the weighting coefficient is that we can identify the relative importance of the models during different sample periods. Waggoner and Zha (2012) show that the DSGE model plays an important role relative to a BVAR model only in the late 1970s and the early 1980s.

We emphasize that we do not assume any regime-switching in the interest rate dynamics under the bond pricing model. Because our estimation is divided into two stages, the Markov-switching of s_t reflects only the prediction densities of the two models as estimated in stage 1, which in turns depend implicitly on the observed dataset Y_T^O. As a consequence, the regime s_t only indicates the particular times of history in which the model dominates in terms of its prediction ability but not the structural change in the interest rate dynamics.

7.4.4 Dynamic prediction pooling

Finally, we adopt the dynamic prediction pooling scheme as proposed in Negro et al. (2014). The idea is to generate a smooth time-varying weighting coefficient, $\lambda_t \in [0, 1]$, based on a probit transformation of a AR(1) process of a latent variable x_t with the autocorrelation coefficient ρ as

$$\begin{aligned} \lambda_t &= N(x_t), \\ x_t &= \rho x_{t-1} + \sqrt{1 - \rho^2}\varepsilon_t, \end{aligned}$$

where $N(\cdot)$ is the cumulative density function of standard normal distribution, the disturbance term follows $\varepsilon_t \sim N(0,1)$ and the initial value takes $x_0 \sim N(0,1)$. The autocorrelation coefficient ρ captures how smooth the weighting coefficient can change over time. When $\rho = 1$, the model reduces to the case of static prediction pooling in Geweke and Amisano (2011) by taking $\lambda_t = \lambda$. When $\rho = 0$, it indicates that λ_t are serially-independent and follows a random walk. Denote $\Lambda_T = (\lambda_1, ..., \lambda_T)$, we have the log score of dynamic prediction pooling as

$$LPS^{DP}(\Lambda_T) \equiv \sum_{t=1}^{T} \log\left[\lambda_t\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + (1-\lambda_t)\, p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right)\right]. \tag{7.15}$$

We follow Negro et al. (2014) and fix the parameter ρ as 0.9. Then, we can estimate the time-varying weighting coefficient λ_t as a nonlinear state space model (due to the probit transformation) using particle filtering. We refer the readers to Johannes and Polson (2009) for a survey on the application of particle filtering in financial econometrics.

7.4.5 Comparison with Bayesian model averaging

A related concept of the optimal prediction pooling is the Bayesian model averaging (BMA) method, in which one of the competing models is correctly specified but the econometrician does not know which one is the true model. Under BMA, the predictive distribution for observation y_t^O is obtained by averaging a set of m competing models, that is

$$p^{BMA}\left(y_t^O | Y_{t-1}^O\right) = \sum_{i=1}^{m} p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_i\right) p\left(\mathcal{M}_i | Y_{t-1}^O, \Theta\right),$$

in which $p\left(\mathcal{M}_i | Y_{t-1}^O, \Theta\right)$ is called the posterior model probability or weight. The key idea in BMA is that the posterior model probability is derived based on the marginal likelihood as

$$p\left(\mathcal{M}_k | Y_{t-1}, \Theta\right) = \frac{p^{ML}\left(y_t^O | Y_{t-1}^O, \mathcal{M}_k\right)}{\sum_{i=1}^{m} p^{ML}\left(y_t | Y_{t-1}^O, \mathcal{M}_i\right)},$$

where $p^{ML}(\cdot)$ is given by (7.12). Since the objective of BMA is for model selection, the computation of model probability often produces extreme weighting on a particular model. This is not practically useful when such a weighting is used to do pooled forecasting because some of the models are simply ignored. We refer the readers to Hoeting et al. (1999) for a throughout discussion.

7.4.6 Estimation method

We adopt a two-stage procedure in which the posterior parameters are saved from the individual model estimation, and then are used for the forecasting and computation of the prediction density/score. In the first stage, posterior estimates of parameters, $p(\Theta|Y_{t-1}^O, \mathcal{M})$, under the prediction models, $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2)$, are obtained using the MCMC method based on the full sample period in order to obtain the prediction scores $p\left(y_t^O; Y_{t-1}^O, \mathcal{M}\right)$ of the three methods. For each set of posterior estimate of the prediction macro-finance models $\mathcal{M}$, we then compute the forecast and prediction density by simulation technique in order to make use of the entire joint posterior distribution. To be more specific, we compute the prediction scores as

$$p(y_t^O; Y_{t-1}^O, \mathcal{M}_1) = p\left(y_t^O; Y_{t-1}^O, \Theta_{QTSM}\right),$$

$$p(y_t^O; Y_{t-1}^O, \mathcal{M}_2) = p\left(y_t^O; Y_{t-1}^O, \Theta_{ATSM}\right),$$

in order to evaluate the log-score criteria. In the second stage, we estimate the optimal combination using the log scores from the individual prediction models (ATSM and QTSM) obtained in the previous step, based on the aforementioned three prediction pooling methods, i.e., (7.13), (7.14) and (7.15). We adopt the Bayesian MCMC approach for the static and Markov-switching approaches, and resort to the particle filtering for the dynamic prediction pool with a fixed parameters ρ (see Appendix).

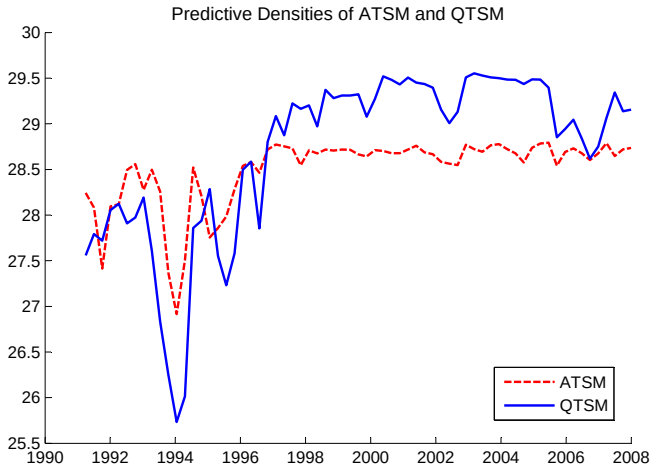
7.5 Estimation Results: Prediction Pooling

7.5.1 Prediction score

We report the combination of the ATSM and QTSM in the prediction of bond yields using the optimal prediction pooling as described in Section 7.4, following Geweke and Amisano (2011), Waggoner and Zha (2012) and Negro et al. (2014). To begin, it is useful to look at the comparison of the predictive densities of the two individual models based on the log-score criteria. While we have performed the comparison using both one-quarter-ahead and four-quarter-ahead forecasts, we only report the charts of four-quarter-ahead forecasts for illustration purpose. As shown in Figure 7.7, the ATSM (dashed red line) dominates the QTSM (solid blue line) for the sampling period from 1990Q1 - 1995Q4 while the QTSM dominates the ATSM when the JGB bond yields are close to zero since 1996Q1. This suggests that one can potentially improve the predictive density by combining appropriately the two models which appear to perform better in

FIGURE 7.7: Log score comparison of the ATSM and QTSM

(A) Four-quarter ahead forecast



Note: The log score at each period is calculated from log $p(y_t^O; Y_{t-1}^O, M_i)$ for i = 1,2, of the individual model such as the QTSM and the ATSM.

different sample periods, i.e., the two models capture different properties of the joint movements of the macro factors and bond yields.

Before we report the results on prediction pooling, let us recall the definition of the log-score function: $LPS(y_t; Y_{t-1}, \text{Pool})$, that is a convex combination of the prediction density at the time-t observation of the ATSM and QTSM as

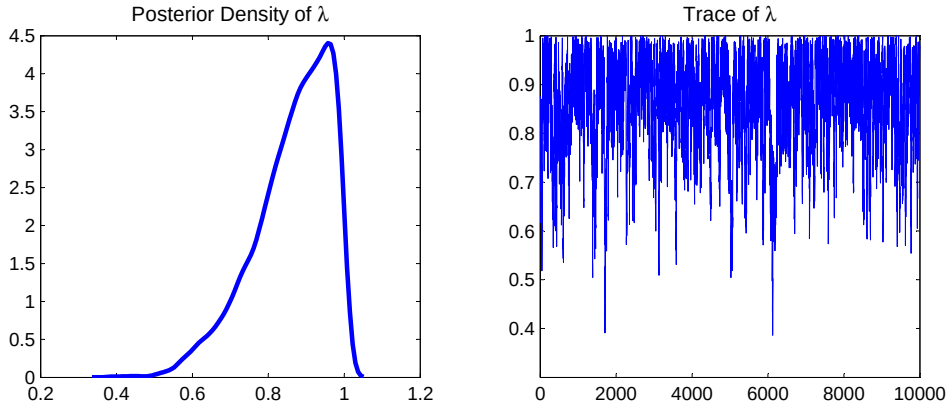
$$LPS(y_t; Y_{t-1}, \text{Pool}) = \log\left[\lambda_t p\left(y_t^O; Y_{t-1}^O, \Theta_{QTSM}\right) + (1-\lambda_t)\, p\left(y_t^O; Y_{t-1}^O, \Theta_{ATSM}\right)\right],$$

in which we take λ_t as the weighting assigned to the QTSM while $1-\lambda_t$ as the weighting assigned to the ATSM, and Θ_{QTSM} and Θ_{ATSM} are the posterior estimates of the QTSM and ATSM parameters over the full sample period, respectively.

7.5.2 Static pooling

Figure 7.8 reports the posterior density and corresponding trace plot of the weighting coefficient λ for the static pooling scheme as (7.13). This coefficient is estimated with MCMC simulation and obtained from 10,000 MCMC draws after discarding the first 5000 draws as burn-in. For the four-quarter-ahead forecast, the posterior distribution of λ is skewed to the left, indicating that the parameter restriction of $\lambda \leq 1$ is binding and one should over-weight the QTSM model and under-weight the ATSM. Table 7.3 reports the posterior mean of the pooling coefficient as $\lambda = 0.8627$ for the four-quarter-ahead

FIGURE 7.8: Static prediction pool (4Q-ahead forecast)



Note: Static pooling model is calculated from (7.13). The weighting coefficient on QTSM, λ, is estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 burn-in draws.

forecast. We also compute the simulation inefficiency statistics as in Kim et al. (1998) and show that MCMC draws of the parameter λ are efficient.

7.5.3 Markov-switching pooling

Figure 7.9 reports the estimation for the Markov-switching pooling scheme, (7.14), in which the extra model parameters include the transition matrix of the Markov chain and the corresponding weighting coefficients λ_1 and λ_2 under the two regimes $s_t = 1$ and $s_t = 2$. Figure 7.9(a) shows the regime variable s_t and the time-varying weight λ_t calculated from s_t.[5] The solid black and red lines denote their posterior means and medians, respectively, while the blue shaded area represents their 90% credible interval. The right panel of Figure 7.9(a) also shows that the posterior density of the MCMC draws of λ_1 and λ_2 are concentrated around 0 and 1 respectively, which means that we can clearly distinguish the two regimes for model combination. Figure 7.9(b) shows the histograms and traces of the MCMC draws of the both parameters λ_1 and λ_2. Again, these posterior estimates are also obtained from MCMC 10,000 draws after discarding the first 5000 draws as burn-in. As shown in Table 7.4, the posterior means of the weighting coefficients are $\lambda_1 = 0.1305$ and $\lambda_2 = 0.9451$ for four-quarter-ahead forecast. This indicates that $s_t = 1$ corresponds to the regime in which the ATSM dominates while $s_t = 2$ corresponds to regime in which the QTSM dominates as the right panel of

[5]The draws of the time-varying weight λ_t are conducted based on the following equation. $\lambda_t = \lambda_1 \times Prob(s_t = 1) + \lambda_2 \times Prob(s_t = 2)$.

TABLE 7.3: Posterior estimates of the static prediction pool

(A) One-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ	0.2982	0.5466	0.0615	0.1480	8.22

(B) Four-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ	0.8627	0.9886	0.6631	0.1014	8.22

Notes:

1. λ denotes the constant weighting coefficient determined in the following optimal prediction pool:

$$p\left(y_t; Y_{t-1}, \mathcal{M}\right) = \lambda p\left(y_t; Y_{t-1}, \mathcal{M}_1\right) + (1-\lambda)\, p\left(y_t; Y_{t-1}, \mathcal{M}_2\right), \qquad 0 \leq \lambda \leq 1,$$

   where $\mathcal{M}_1$ = QTSM $\mathcal{M}_2$ = ATSM, and $p\left(y_t; Y_{t-1}, \mathcal{M}_1\right)$ denotes log prediction score.

2. The coefficient λ is estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 draws. And the posterior means, the standard deviations (Std. Dev.), and the percentiles are derived from the sampled draws.

3. The simulation inefficiency statistic is a useful diagnostic for measuring how well the chain mixes according to Kim, Shephard, Chib (1998). The statistic is derived from:

$$\hat{R}_{B_M} = 1 + \frac{2B_M}{B_M - 1} \sum_{i=1}^{B_M} K(\frac{i}{B_M})\hat{\rho}(i),$$

   where $\hat{p}(i)$ is an estimate of the autocorrelation at lag i of MCMC sampler, B_M represents the bandwidth and K the Parzen Kernel.

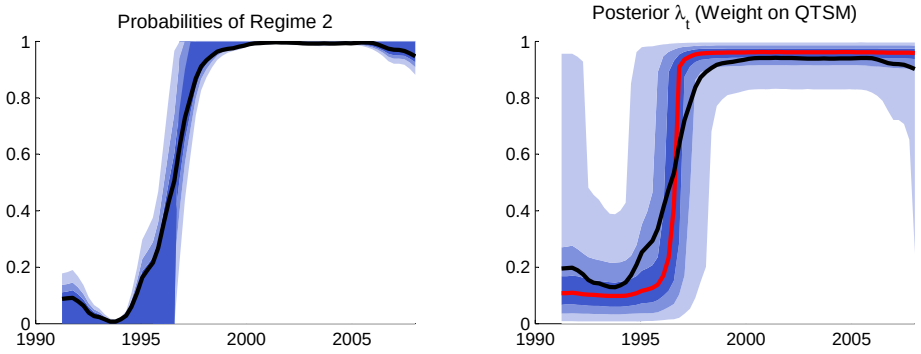
Figure 7.9(a). We can also look at the conditional probability for a regime $j = 1, 2$ as

$$P\left(s_t = j \middle| S_{-t}, Y_T\right) = \frac{g\left(s_t = j \middle| S_{-t}, Y_T\right)}{\sum_{j=1}^{2} g\left(s_t = j \middle| S_{-t}, Y_T\right)},$$

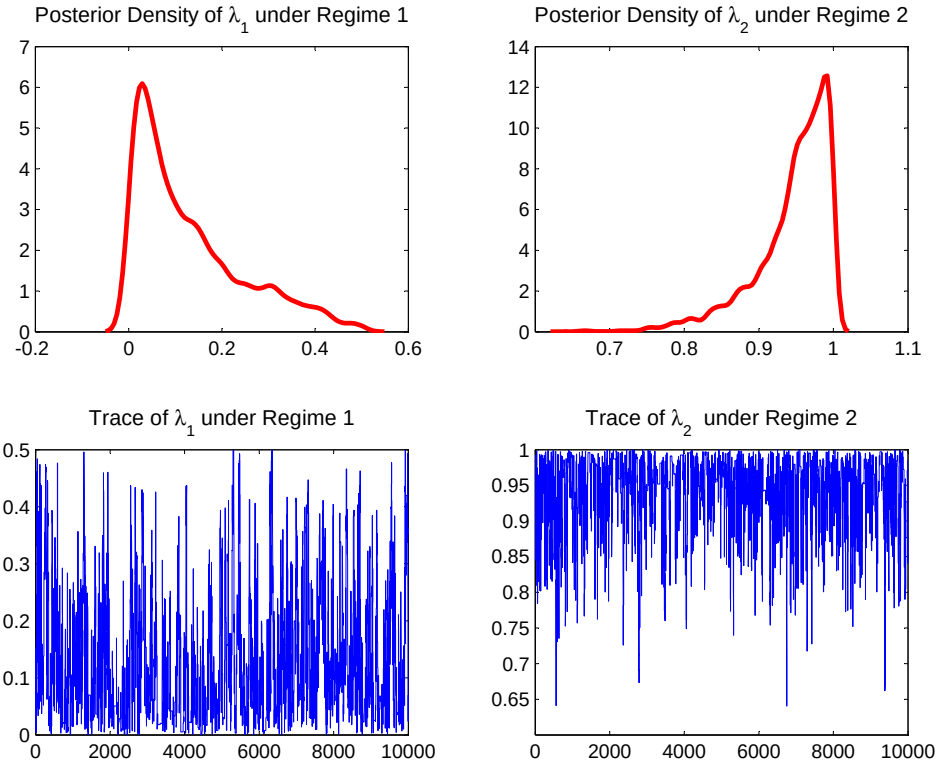
where $g\left(s_t \middle| S_{-t}, Y_T\right)$ is the posterior distribution of s_t related to the pooled prediction density (see Appendix). The left panel of Figure 9(a) shows that the probability of regime 2 is close to one since 1996 as the JGB yields continued to move towards the zero lower bound amid the Bank of Japan's interest rate cuts. In particular, it is interesting to look closer to the posterior distribution of λ_1: as the model switches gradually from the ATSM to QTSM in the early 1990s, the dispersion of λ_1 is large, indicating that both the two models might useful in explaining the bond yields; as the Bank of Japan adopts the ZIRP from 1999Q1 to 2000Q3, the distribution of λ_1 is concentrated around 1, indicating that the QTSM model captures much better the joint dynamics of bond yields and the macroeconomy under the ZIRP.

FIGURE 7.9: Markov-switching prediction pool (4Q-ahead forecast)

(A) Probability of regime 2 (QTSM)



(B) Posterior distributions



Notes:

1. Markov-switching pooling model is calculated from (7.14). The weighting coefficients on QTSM, λ_i, are estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 burn-in draws.
2. The solid black and red lines denote their posterior means and medians, respectively, and the blue shaded area represents their 90% confidence interval.

TABLE 7.4: Posterior estimates of the Markov-switching prediction pool

(A) One-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ_1	0.0899	0.2670	0.0039	0.0831	8.22
λ_2	0.9014	0.9951	0.7142	0.0906	15.44

(B) Four-quarter-ahead forecast

Parameter	Mean	90 Percentile	10 Percentile	Std. Dev.	Inefficiency
λ_1	0.1305	0.3733	0.0070	0.1167	8.22
λ_2	0.9451	0.9970	0.8395	0.0507	15.44

Notes:

1. λ_1 and λ_2 denote the regime switching weighting coefficient determined by regime variables s_t in the following optimal prediction pool:

$$p\left(y_t;\, Y_{t-1}, \mathcal{M},\, s_t\right) = \lambda\left(s_t\right) p\left(y_t; Y_{t-1}, \mathcal{M}_1\right) + \left(1 - \lambda\left(s_t\right)\right) p\left(y_t; Y_{t-1}, \mathcal{M}_2\right),$$

   where $\mathcal{M}_1 = \text{QTSM}, \mathcal{M}_2 = \text{ATSM}$ and

$$\lambda_t = \lambda\left(s_t\right) = \begin{cases} \lambda_1, & s_t = 1 \\ \lambda_2, & s_t = 2 \end{cases},$$

2. The coefficient λ is estimated with MCMC simulation and obtained from 10,000 draws after discarding the first 5000 draws, and the posterior means, the standard deviations (Std. Dev.), and the percentiles are derived from the sampled draws.
3. The simulation inefficiency statistic is a useful diagnostic for measuring how well the chain mixes according to Kim, Shephard, Chib (1998). The statistic is derived from:

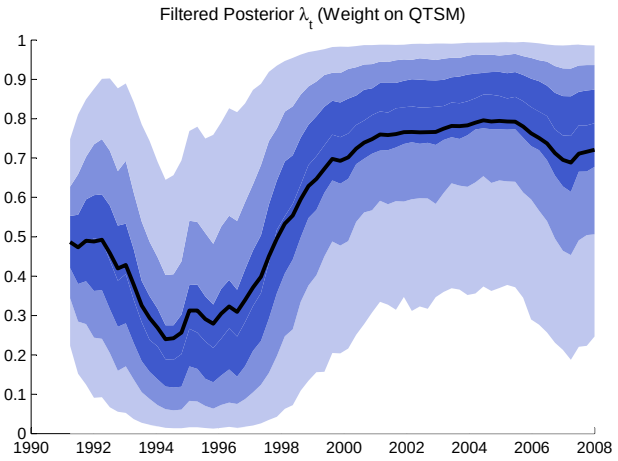
$$\hat{R}_{B_M} = 1 + \frac{2B_M}{B_M - 1} \sum_{i=1}^{B_M} K(\frac{i}{B_M})\hat{\rho}(i),$$

   where $\hat{p}(i)$ is an estimate of the autocorrelation at lag i of MCMC sampler, B_M represents the bandwidth and K the Parzen Kernel.

7.5.4 Dynamic pooling

Figure 7.10 shows the estimation for the dynamic pooling scheme which imposes a smooth transition in between the two selected models as described in (7.15). The solid black line denotes the posterior means of the time varying weighting coefficients while the blue shaded area represents their 90% credible interval. The estimation is obtained from 5000 draws of particle filter with constant autocorrelation coefficient ρ fixed as 0.9 following Negro et al. (2014), which constraints the time-variation of the weighting to be relatively slow-moving. Similar to the Markov-switching pooling, the weighting to QTSM increases after 1995Q1 and keep dominating the ATSM afterwards. However, the dispersion of the posterior distribution of λ_t appears to be large with the median fluctuating around 0.3 to 0.8 for the four-quarter-ahead forecasts. In contrast to the Markov-switching pooling, the dynamic pooling scheme does not allow us to obtain a clear cut in between QTSM and ATSM.

FIGURE 7.10: Dynamic prediction pool (4Q-ahead forecast)



Notes:

1. Dynamic pooling model is calculated from ((7.15)). The time-varying coefficient is obtained from 5000 draws of particle filter with constant autocorrelation coefficient ρ fixed as 0.9 following Del Negro et al (2013).
2. The solid black line denotes their posterior means and the blue shaded area represents their 90% confidence interval.

7.5.5 Comparison

Lastly, let us compare the performance of different models and pooling schemes in terms of the log score of prediction density. Table 7.5 summarizes the corresponding cumulative log score performance for the one-quarter-ahead and four-quarter-ahead forecasts. Figure 11 shows the time series of the log score of the three pooling schemes as well as those of the two individual models. As can be seen in Table 7.5, the Markov-switching (MS) pooling scheme produces the best cumulative log score: this is because it allows one to combine the ATSM and QTSM efficiently by switching from the ATSM before 1995 to the QTSM after 1995 as depicted in Figure 7.11. Interestingly, the static pooling scheme only marginally improve the cumulative log-score performance for the four-quarter-ahead forecast, although it may improve the prediction density in certain sub-sample periods. On the other hand, the two time-varying pooling schemes produce a better improvement over individual models. In particular, the Markov-switching pooling scheme appears to be more effective than the dynamic pooling scheme in capturing the abrupt change in bond yield movements when the monetary policy switches at late 1995. This suggests that an appropriate pooling scheme is important for one to achieve an overall improvement (in terms of the log-score criteria) in the prediction of future bond yields when models are combined.

TABLE 7.5: Cumulative log scores

(A) One-quarter-ahead forecast

Component Models		Model Pooling	
Model	Log Score	Methods	Log Score
ATSM	2006.85	Static	2008.38
QTSM	1987.93	Markov Switching	2018.13
		Dynamic	2013.82

(B) Four-quarter-ahead forecast

Component Models		Model Pooling	
Model	Log Score	Methods	Log Score
ATSM	1909.20	Static	1922.78
QTSM	1923.03	Markov Switching	1930.92
		Dynamic	1925.87

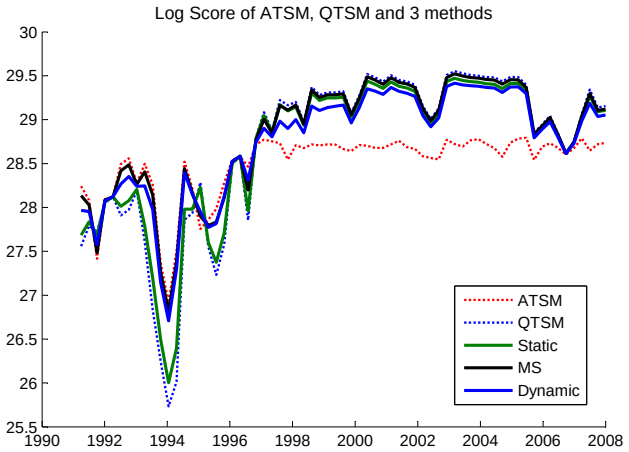
Notes:

1. The predictive densities for the ATSM and QTSM are obtained by simulation using the MCMC draws of the posterior model parameters.
2. The cumulative log score is computed as

$$\sum_{t=1}^{T} \log\left[\lambda_t p\left(y_t; Y_{t-1}, \mathcal{M}_1\right) + \left(1-\lambda_t\right) p\left(y_t; Y_{t-1}, \mathcal{M}_2\right)\right]$$

where $\mathcal{M}_1 = \text{QTSM}, \mathcal{M}_2 = \text{ATSM}$.

FIGURE 7.11: Log score comparison (4Q-ahead forecast)



Note: The log scores at each period is calculated from log $p(y_t^O; Y_{t-1}^O, M_i)$ for i = 1,2, of the individual model such as the QTSM and the ATSM.

7.6 Summary

In this chapter, we study the optimal prediction pool of the Gaussian ATSM and QTSM with macro-finance features using the JGB data from 1990Q1 to 2008Q3 that cover the zero interest rate policy in Japan. Our estimation results show that the QTSM provides a more realistic description of bond yields when the zero lower bound is binding, although the ATSM appears to provide a better fit to bond yields and macroeconomic variables simultaneously. This suggests that one should combine the two models for the prediction of future bond yields under different market scenarios. In particular, we find that the Markov-switching pooling is the most effective in capturing the abrupt change in bond yield movements as the monetary policy switches.

For future research, it is instructive to explore a wider combination of macroeconomic variables, such as unemployment rate, M2 growth and credit-to-GDP ratio. Moreover, it is interesting to repeat the exercise using the US treasury yield data since the financial crisis of 2008, although the history may be limited for a robust statistical identification. A potential remedy is to use macroeconomic variables with higher frequency such as monthly data. An alternative is to use the estimation technique with mixing frequency data such as the one proposed in Camacho and Perez-Quiros (2010).

7.7 Appendix

7.7.1 Bond pricing

For notational convenience, we will take $\mu^Q = \mu$ and $\Phi^Q = \Phi$ as the risk-neutral parameters and all expectations are under the risk neutral measure Q.

ATSM

The n-period zero coupon bond price can be formulated as

$$\begin{aligned} P_t^n &= \mathbb{E}_t\left[e^{-r_t} P_{t+1}^{n-1}\right] \\ &= \mathbb{E}_t\left[\exp\left(-r_t + A_{n-1} + B_{n-1}^T X_{t+1}\right)\right], \end{aligned}$$

where $r_t = \delta_0 + \delta_1^T X_t$ and X_t follows the VAR dynamics $X_{t+1} = \mu + \Phi X_t + \Sigma\varepsilon_{t+1}$ with $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$. We can substitute the expression of X_{t+1} such that

$$\begin{aligned} P_t^n &= \mathbb{E}_t\left[\exp\left(-r_t + A_{n-1} + B_{n-1}^T X_{t+1}\right)\right] \\ &= \exp\left(-r_t + A_{n-1} + \mu + B_{n-1}^T \Phi X_t\right) \mathbb{E}_t\left[\exp\left(B_{n-1}^T \Sigma\varepsilon_{t+1}\right)\right]. \end{aligned}$$

Then, we can make use of the moment generating function of $\varepsilon \sim N(\mathbf{0}, \mathbf{I})$ to compute the expectation as

$$\mathbb{E}_t\left[\exp\left(B_{n-1}^T \Sigma \varepsilon\right)\right] = \exp\left[\frac{1}{2} B_{n-1} \Sigma \Sigma^T B_{n-1}^T\right].$$

by collecting separately the constant terms and linear terms in X_t, we obtain the recursive relationship for ATSM.

QTSM

The n-period zero coupon bond price can be formulated as

$$\begin{aligned} P_t^n &= \mathbb{E}_t\left[e^{-r_t} P_{t+1}^{n-1}\right] \\ &= \mathbb{E}_t\left[\exp\left(-r_t + A_{n-1} + B_{n-1}^T X_{t+1} + X_{t+1}^T C_{n-1} X_{t+1}\right)\right], \end{aligned}$$

where $r_t = \alpha_0 + \beta_0^T X_t + X_t^T \Psi_0 X_t$ and X_t follows the VAR dynamics $X_{t+1} = \mu + \Phi X_t + \Sigma \varepsilon_{t+1}$ with $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$. Similarly, we substitute the expression of X_{t+1} such that

$$\left(\mu + \Phi X_t + \Sigma \varepsilon_{t+1}\right)^T C_{n-1} \left(\mu + \Phi X_t + \Sigma \varepsilon_{t+1}\right) = 2\left(\mu + \Phi X_t\right)^T C_{n-1} \Sigma \varepsilon_{t+1},$$

and hence

$$\begin{aligned} P_t^n &= \exp\left(-r_t + A_{n-1} + B_{n-1}^T \mu + \Phi X_t + \left(\mu + \Phi X_t\right)^T C_{n-1}\left(\mu + \Phi X_t\right)\right) \\ &\quad \times \mathbb{E}_t\left[\exp\left(\Gamma_0^T \varepsilon_{t+1} + \varepsilon_{t+1}^T \Gamma_1 \varepsilon_{t+1}\right)\right], \end{aligned}$$

where

$$\Gamma_0^T = B_{n-1}^T \Sigma + 2\left(\mu + \Phi X_t\right)^T C_{n-1} \Sigma, \qquad \Gamma_1 = \Sigma^T C_{n-1} \Sigma.$$

In this case, we can make use of the (exponential) quadratic-form expectation for $\varepsilon \sim N(\mathbf{0}, \mathbf{I})$ as

$$\mathbb{E}_t\left[\exp\left(\Gamma_0^T \varepsilon + \varepsilon^T \Gamma_1 \varepsilon\right)\right] = \exp\left[-\frac{1}{2}\det\left(\mathbf{I} - 2\Gamma_1\right) + \frac{1}{2}\Gamma_0\left(\mathbf{I} - 2\Gamma_1\right)^{-1}\Gamma_0\right].$$

See, for example, Chapter 12 in Andersen and Piterberg (2010). Therefore,

$$\begin{aligned} &\mathbb{E}_t\left[\exp\left(\Gamma_0^T \varepsilon_{t+1} + \varepsilon_{t+1}^T \Gamma_1 \varepsilon_{t+1}\right)\right] \\ = &\exp\left(-\frac{1}{2}\det\left(\mathbf{I} - 2\Sigma^T C_{n-1} \Sigma\right)\right) \\ &\times \exp\left(\left(B_{n-1}^T \Sigma + 2\left(\mu + \Phi X_t\right)^T C_{n-1} \Sigma\right)\left(\mathbf{I} - 2\Sigma^T C_{n-1}\Sigma\right)^{-1}\right. \\ &\left.\left(B_{n-1}^T \Sigma + 2\left(\mu + \Phi X_t\right)^T C_{n-1}\Sigma\right)^T\right), \end{aligned}$$

collecting separately the constant terms, linear terms in X_t and quadratic terms in X_t, we obtain the recursive relationship for QTSM. We calibrate the initial loadings of QTSM model as: $\alpha_0 = 0$, $\beta_0 = 0$ and $\Psi_0 = diag(50, 0, 0)$ in order to match with the range of the bond yields data.

7.7.2 Bayesian estimation

State space formulation

We discuss the Bayesian estimation procedure in more details. First-of-all, it is useful to re-write the state space model in Section 3.4 as follows:

- Measurement equation. Factor loadings a_n, b_n, and c_n are derived from the recursive relationship as described in Section 3.3. The measurement equations for the observable bond yields $\hat{y}_t^n$ and macro factors $\hat{X}_t$ are related to the latent factors X_t as

$$\hat{X}_t = X_t + \omega_{X,t},$$

and

$$\hat{y}_t^n = a_n + b_n^T X_t + X_t^T c_n X_t + \omega_{n,t}.$$

- State equation. The state equation with the parameters μ^P and Φ^P is given by

$$X_{t+1} = \mu^P + \Phi^P X_t + \Sigma\varepsilon_{t+1},$$

which is a standard VAR(1) system.

MCMC algorithm

In the estimation, we need to tackle a non-linear state space model due to the complicated form of the measurement equations. By following the procedure in Ang et al. (2011), we adopt Metropolis-Hastings (MH) within Gibbs with a single-move sampler for the unobservable variables $X_t = (L_t, g_t, \pi_t)$. For notational convenience, we denote m to be the m-th draw or m-th iteration of the MCMC procedure. Moreover, we use $\varepsilon^m \sim N(0, s^2)$ to denote a draw from a i.i.d. normal with a variance s^2, which is tuned for each random-walk MH sampling procedure employed in various steps.

The algorithm of MCMC is consisted of the following five steps. Note that step 2 and step 3 are implemented as a standard Bayesian VAR with Minnesota prior.

- **Step 1: Drawing the latent factor $X_t = (L_t, g_t, \pi_t)$.** We adopt the single-move sampler and generate the latent factors using random walk MH with the

conditional posterior density:

$$P(X_t|X_{t-1},\tilde{Y},\Theta) \propto P(X_t|X_{t-1})P(\tilde{Y}_t|\, X_t,\, \Theta)P(X_{t+1}|X_t)$$

where

$$P(X_t|X_{t-1},\Theta) \propto \exp(-\frac{1}{2}(X_t - \mu^P - \Phi^P X_{t-1})^T(\Sigma\Sigma^T)^{-1}(X_t - \mu^P - \Phi^P X_{t-1})\,)$$

and

$$P(\widetilde{Y}_t|X_t,\Theta) \propto \left(-\frac{1}{2}\sum_n \left[\frac{\left(\widetilde{y}_t^n - \left(a_n + b_n^T X_t + X_t^T c_n X_t\right)\right)^2}{\sigma_n^2}\right]\right)$$

where $\tilde{Y}_t$ is observable variables including yields and macro variables and Θ collects the parameters. The standard deviation of the random walk MH step is taken to be 0.0001 (i.e., 1 bps).

- **Step 2: Drawing μ^P and Φ^p under the real-world measure P.** We use the Gibbs sampler to draw μ^P and Φ^p with the conditional posterior density

$$P(\mu^P,\Phi^P|\Theta_-,\, X,\, \tilde{Y}) \propto P(X|\mu^P,\Phi^P,\Sigma)P(\mu^P,\Phi^P)$$

where $P(X|\mu^P,\Phi^P,\Sigma)$ is the likelihood function and $P(\mu^P,\Phi^P)$ is the prior as in Negro and Schorfheide (2010).

- **Step 3: Drawing $\Sigma\Sigma'$, the variance of state equation.** We take the inverse Wishart distribution as the prior and sample from the proposal density

$$q(\Sigma\Sigma') = P(X|\mu,\Phi,\Sigma)P(\Sigma\Sigma'),$$

where $P(X|\mu,\Phi,\Sigma)$ and $P(\Sigma\Sigma')$ are the likelihood function and prior, respectively.

- **Step 4: Drawing $\mu^Q$and Φ^Q under the risk-neutral measure Q.** We use the random walk MH algorithm and sample $\mu^Q$and Φ^Q from a proposal draw using the random walk process $x^{m+1} = x^m + \varepsilon^m$. The proposal draw is then accepted with the probability

$$\alpha = \min\left\{\frac{P(\tilde{Y}|(\mu^Q,\Phi^Q)^{m+1},\Theta_-,X)}{P(\tilde{Y}|(\mu^Q,\Phi^Q)^m,\Theta_-,X)},\, 1\right\},$$

where $P(\tilde{Y}|(\mu^Q,\Phi^Q)^{m+1},\Theta_-,X)$ is the likelihood function or the posterior density as we assume a flat prior as in Ang et al. (2011). The standard deviation of the random walk MH step is taken to be 0.1% of the magnitude of the initial parameters.

- **Step 5: Drawing the variance of measurement error (σ_u).** We take the inverted Gamma distribution as prior with $IG(0, 0.0025^2)$ in order to sample σ_u.

We note that it is possible to estimate the non-linear state space model using the particle filter as in Andreasen et al. (2013). However, we find that one needs to spend an extensive computational time to estimate our 3-factor QTSM when the filter is used along with the Bayesian estimation. The comparison of the two approaches are left for further research.

Optimal pooling

Next, we discuss the estimation procedures as related to the three optimal pooling schemes. The main idea is that we can treat the pooled prediction density as the likelihood of the corresponding model parameters for the proposed pooling scheme. Then, we can derive and sample the posterior distributions of relevant parameters based on standard Bayesian MCMC techniques.

Static prediction pooling

To estimate the constant weighting λ, we adopt the random walk MH algorithm to sample its posterior distribution with a flat prior. To be specific, we generate a proposal draw $\lambda^{m+1} = \lambda^m + \varepsilon^m$, compute the likelihood

$$p(\lambda^{m+1}) = \sum_{t=1}^{T} \log \left[\lambda^{m+1} p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + \left(1 - \lambda^{m+1}\right) p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right) \right],$$

and accept the draw with the probability

$$\alpha = \min \left\{ \frac{p(\lambda^{m+1}))}{p(\lambda^m)}, 1 \right\}.$$

Markov-switching prediction pooling

By following by Albert and Chib (1993), we use a MH within Gibbs algorithm consisted of the following three steps.

1. The regime s_t at period t is sampled by a single-move algorithm with backward recursion;
2. The regime-dependent weighting λ_j for $j = 1, 2$ are sampled in a block by random-walk MH;
3. The transitional probabilities q_{11} and q_{22} are sampled directly using conjugate priors (beta distribution).

Denote $S_{-t} = (s_1, ..., s_{t-1}, s_{t+1}, ..., s_T)$ and $Y_T = (y_1, ..., y_T)$. The single-move algorithm draws the regime state based on a backward recursion for $t = T, T-1, ..., 1$ with the conditional probability

$$P\left(s_t = j \middle| S_{-t}, Y_T\right) = \frac{g\left(s_t = j \middle| S_{-t}, Y_T\right)}{\sum_{j=1}^{2} g\left(s_t = j \middle| S_{-t}, Y_T\right)},$$

where $g\left(\cdot \middle| S_{-t}, Y_T\right)$ is the posterior distribution of s_t given by

$$g\left(s_t \middle| S_{-t}, Y_T\right) \propto p\left(s_{t+1} \middle| s_t\right) f\left(y_t^O \middle| s_t\right) p\left(s_t \middle| s_{t-1}\right),$$

in which

$$f\left(y_t^O \middle| s_t\right) = \lambda_{s_t}^m p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + \left(1 - \lambda_{s_t}^m\right) p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right),$$

is the pooled prediction density at regime s_t, and

$$p\left(s_{t+1} \middle| s_t\right) \quad \text{and} \quad p\left(s_t \middle| s_{t-1}\right)$$

are the transitional probabilities of the Markov chain. Since s_t is a Bernoulli random variable, we can easily draw s_t and accept it with the probability $p\left(s_t = j \middle| S_{-t}, Y_T\right)$.

In step 2, the corresponding weighting λ_1^m and λ_2^m are sampled in a block based on the random-walk MH procedure as

$$\begin{pmatrix} \lambda_1^{m+1} \\ \lambda_2^{m+1} \end{pmatrix} = \begin{pmatrix} \lambda_1^m \\ \lambda_2^m \end{pmatrix} + \begin{pmatrix} \varepsilon_1^m \\ \varepsilon_2^m \end{pmatrix}, \quad \epsilon_i^m \sim \text{i.i.d. } N(0, s^2),$$

with the likelihood

$$p(\lambda_1^{m+1}, \lambda_2^{m+1}) = \sum_{t=1}^{T} \log \left\{ \sum_{j=1}^{2} \left[\lambda_j^{m+1} p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + \left(1 - \lambda_j^{m+1}\right) p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right) \right] \right\}.$$

We then accept the draw with the probability

$$\alpha = \min \left\{ \frac{p(\lambda_1^{m+1}, \lambda_2^{m+1})}{p(\lambda_1^m, \lambda_2^m)}, 1 \right\}.$$

In step 3, the transitional probabilities can be sampled by using the beta distribution as conjugate prior such that (Albert and Chib, 1993)

$$\begin{aligned} q_{11} \middle| s_{1:T} &\sim \beta\left(u_{11} + n_{11}, u_{12} + n_{12}\right), \\ q_{22} \middle| s_{1:T} &\sim \beta\left(u_{22} + n_{22}, u_{21} + n_{21}\right), \end{aligned}$$

where $\beta(\alpha_0, \alpha_1) \propto z^{\alpha_0 - 1} (1-z)^{\alpha_1 - 1}$ is the beta distribution and n_{ij} is the number of transition from state i to j as counted from the samples of $s_{1:T}$. We set the prior hyper-parameters $u_{11} = 9$, $u_{12} = 1$, $u_{22} = 9$ and $u_{21} = 1$, indicating that the prior means of q_{11} and q_{11} are $\alpha_0/(\alpha_0 + \alpha_1) = 0.1$, and hence a transition is relatively rare.

Dynamic prediction pooling

We follow closely the procedure in Del Negro et al. (2014) to estimate the dynamic prediction pooling. In particular, we implement the particle filter for $\{x_t\}_{1:T}$ based on the standard sampling importance resampling (SIR) algorithm:

1. To initialize, we sample the particles x_1^i for $i = 1, 2, ..., I$ from i.i.d. $N(0, s^2)$ and assign the equal weighting $\omega_1^i = 1$.

2. For $t = 2, 3, .., T$, propagate of particles by the AR(1) dynamics

$$x_t^i = \rho x_{t-1}^i + \sqrt{1-\rho^2}\varepsilon_t^i, \quad \varepsilon_t^i \sim N(0, s^2),$$

   and obtain the sample $\lambda_t^i = N\left(x_t^i\right)$ based on the probit transformation.

3. Compute the likelihood using the pooled prediction density as

$$\omega_t^i = \lambda_t^i p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_1\right) + \left(1 - \lambda_t^i\right) p\left(y_t^O; Y_{t-1}^O, \mathcal{M}_2\right),$$

   and then compute normalized likelihood (weighting) as $\omega_t^i = \omega_t^i / \sum_{i=1}^I \omega_t^i$.

4. Resample the particles x_t^i by using the normalized weighting ω_t^i and then return to Step 2.

We set the autocorrelation coefficient ρ to be a constant as $\rho = 0.9$ and employ 5,000 particles to achieve a good approximation to the true distribution.

Chapter 8

Conclusion

In this thesis, I address several important issues of the financial market in the aftermath of the Lehman crisis from the perspective in financial engineering, including the decision making for a trading strategy, the pricing of financial instruments and the empirical analysis of bond yields data near the zero lower bound. The topics are motivated by the new practical and theoretical challenges as emerged from the recent financial crisis of 2008.

In the first part, I study the optimal stopping problem related to a short-selling strategy and derive the optimal timing for a short-covering under the setting of geometric Brownian motion and regime-switching stock price process. I show that the type of the stopping rule depends on the benefit and cost of holding the short position and I conduct detailed comparative analysis on the model parameters. For practical application, I calibrate the regime-switching model to actual data of the Nikkei 225 stock market and highlight the importance of developing a regime-dependent stopping rule.

In the second part, I investigate several problems related to the pricing of financial instruments, including European options, capped variance swaps and contingent convertibles. I adopt an equity-credit hybrid modeling approach and study the pricing of capped variance swaps in the presence of a jump-to-default feature. The pricing formula are obtained in closed-form for continuous-monitoring and in terms of an approximation for discrete-monitoring. I then extend the equity-credit hybrid modeling approach to the pricing of contingent convertible bond, which is a relatively new financial instrument emerging after the Lehman crisis. An integral equation approach is employed to tackle the first-passage-time problem for a bivariate Gaussian process as related to the pricing of contingent convertibles. Furthermore, I study the asymptotic expansion approach for the multifactor stochastic volatility model of Heston (1993) and perform extensive calibration exercise to the market data of the S&P 500 and Nikkei 225 option markets.

In the third part, I conduct the empirical study of non-linear term structure model with macro-finance feature using the Bayesian MCMC method. I study the model performances of the Gaussian affine and quadratic term structure models under the zero interest rate policy of Japan since the early 1990s. I investigate the pooled prediction of bond yields using a log-score criteria and compare the empirical performance of affine and quadratic term structure models. The result shows that the quadratic model provides a more realistic description of bond yields when interest rates are close to the zero lower bound.

With the strong focus on practical relevancy, I hope that the mathematical models and analysis developed in this thesis will be useful not only for academic researchers in financial engineering, but also for market participants and policy makers in the banking and finance industry.

Bibliography

[1] Acharya, V. and Johnson, T. (2007). Insider trading in credit derivatives. *Journal of Financial Economics*, 84:110–141.

[2] Adrian, T., Begalle, B., Copeland, A., and Martin, A. (2013). Repo and securities lending. Federal Reserve Bank of New York Staff Reports, no. 529.

[3] Ahn, D., Dittmar, R., and Gallant, A. (2002). Quadratic term structure models: Theory and evidence. *Review of Financial Studies*, 15:243–288.

[4] Albert, J. and Chib, S. (1993). Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business and Economic Statistics*, 11(1):1–15.

[5] Albul, B., Jaffee, D., and Tchistyi, A. (2010). Contingent convertible bonds and capital structure decisions. University of California Berkeley Coleman Fung Risk Management Research Center Working Paper 2010-01.

[6] Alexander, C. and Leontsinis, S. (2012). Model risk in variance swap rates. Working Paper, University of Reading.

[7] Alós, E., Leòn, J. A., and Vives, J. (2007). On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4):571–589.

[8] Andersen, L. and Piterberg, V. (2010). *Interest Rate Modeling. Volume 2: Term Structure Models*. Atlantic Financial Press.

[9] Andreasen, M. and Meldrum, A. (2013). Likelihood inference in non-linear term structure models: the importance of the lower bound. Bank of England Working Paper No. 481.

[10] Ang, A., Boivin, J., Dong, S., and Loo-Kung, R. (2011). Monetary policy shifts and the term structure. *Review of Economic Studies*, 78:429–457.

[11] Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50(4):745–787.

[12] Ang, A., Piazzesi, M., and Wei, M. (2006). What does the yield curve tell us about gdp growth? *Journal of Econometrics*, 131:745–787.

[13] Antonelli, F. and Scarlatti, S. (2009). Pricing options under stochastic volatility: a power series approach. *Finance and Stochastics*, 13(2):269–303.

[14] Asmussen, S. (2003). *Applied Probabilities and Queues*. Springer.

[15] Avdjiev, S., Kartasheva, A., and Bogdanova, B. (2013). Cocos: a primer. *BIS Quarterly Review*.

[16] Baba, N. (2006). Financial market functioning and monetary policy: Japan's experience. *Monetary and Economic Studies*, 24:39–71.

[17] Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049.

[18] Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9(1):69–107.

[19] Battalio, R., Mehran, H., and Schultz, P. (2011). Market declines: Is banning short selling the solution? Federal Reserve Bank of New York Staff Reports, no. 518.

[20] Battalio, R. and Schultz, P. (2011). Regulatory uncertainty and market liquidity: The 2008 short sale ban's impact on equity option markets. *Journal of Finance*, 66(6):2013–2053.

[21] Bayraktar, E. and Yang, B. (2011). Unified framework for pricing credit and equity derivatives. *Mathematical Finance*, 21(3):493–517.

[22] Benhamou, E., Gobet, E., and Miri, M. (2010). Time dependent heston model. *SIAM Journal on Financial Mathematics*, 1(1):289–325.

[23] Bergomi, L. (2008). Dynamic properties of smile models. In Cont, R., editor, *Frontiers in Quantitative Finance: Volatility and Credit Risk Modeling*. John Wiley & Sons, USA.

[24] Bernanke, B., Reinhart, V., and Sack, B. (2004). Monetary policy alternatives at the zero bound: An empirical assessment. *Brookings Papers on Economic Activity*, 2:1–100.

[25] Bernard, C., Courtois, Q. L., and Quittart-Pinon, F. (2008). Pricing derivatives with barriers in a stochastic interest rate environment. *Journal of Economic Dynamics and Control*, 32:2903–2938.

[26] Black, F. and Cox, J. (1976). Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance*, 31:351–367.

[27] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654.

[28] Boehmer, E. and Wu, J. (2013). Short selling and the price discovery process. *Review of Financial Studies*, 26(2):287–322.

[29] Borodin, A. N. and Salminen, P. (2002). *Handbook of Brownian Motion*. Birkhauser, Boston, MA.

[30] Brigo, D., Garcia, J., and Pede, N. (2013). Coco bonds valuation with equity- and credit-calibrated first passage structural models. *Journal of Banking and Finance*.

[31] Brunnermeier, M. (2010). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives*, 23:77–100.

[32] Buehler, H. (2006). Consistent variance curve model. *Finance and Stochastics*, 10(2):178–203.

[33] Buehler, H. (2010). Volatility and dividends - volatility modelling with cash dividends and simple credit risk. Working Paper.

[34] Camacho, M. and Perez-Quiros, G. (2010). Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics*, 25:663–694.

[35] Campi, L., Polbennikov, S., and Sbuelz, A. (2009). Systematic equity-based credit risk: a cev model with jump to default. *Journal of Economic Dynamics and Control*, 33:93–108.

[36] Cao, C., Yu, F., and Zhong, Z. (2010). The information content of option-implied volatility for credit default swap valuation. *Journal of Financial Markets*, 13(3):321–343.

[37] Carr, P. and Linetsky, V. (2006). A jump to default extended cev model: An application of bessel processes. *Finance and Stochastics*, 10:303–330.

[38] Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2(4):61–73.

[39] Carr, P. and Madan, D. (2010). Local volatility enhanced by a jump to default. *SIAM Journal on Financial Mathematics*, 1(1):2–15.

[40] Carr, P. and Wu, L. (2009). Stock options and credit default swaps: A joint framework for valuation and estimation. *Journal of Financial Econometrics*, 8(4):1–41.

[41] Chen, T. (2012). Volatility swap pricing with term heston. Bloomberg Technical Note.

[42] Cheridito, P. and Wugalter, A. (2012). Pricing and hedging in affine models with possibility of default. *SIAM Journal on Financial Mathematics*, 3(1):328–350.

[43] Cheridito, P. and Xu, Z. (2014). A reduced form coco model with deterministic conversion intensity. *Journal of Risk*.

[44] Christoffersen, P., Heston, S., and Jacobs, K. (2009). The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55(12):1914–1932.

[45] Chung, T. (2015). Optimal short-covering with regime switching. Recent Advances in Financial Engineering 2014, editied by Massaki Kijima, Yukio Muromachi and Takashi Shibata.

[46] Chung, T. and Kwok, Y. (2014). Equity-credit modeling under affine jump-diffusion models with jump-to-default. *Journal of Financial Engineering*, 1(2).

[47] Chung, T. and Tanaka, K. (2015). Optimal timing for short-covering of an illiquid security. *Journal of the Operations Research Society of Japan*, 58(2).

[48] Coculescu, D., Geman, H., and Jeanblanc, M. (2008). Valuation of default-sensitive claims under imperfect information. *Finance and Stochastic*, 12:195–218.

[49] Collin-Dufresne, P. and Goldstein, R. (2001). Do credit spreads reflect stationary leverage ratios? *Journal of Finance*, 56(5):1929 – 1957.

[50] Cont, R. and Tankov, P. (2004). Non-parametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance*, 7(3):1–50.

[51] Corcuera, J. M., Spiegeleer, J. D., Ferreiro-Castilla, A., Kyprianou, A., Madan, D., and Schoutens, W. (2013). Pricing of contingent convertibles under smile conform models. *Journal of Credit Risk*, 9(3):121–140.

[52] Cox, J., Ingersoll, J., and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–408.

[53] Cremers, M., Driessen, J., Maenhout, P., and Weinbaum, D. (2008). Individual stock option prices and credit spreads. *Journal of Banking and Finance*, 32:2706–2715.

[54] D'Avolio, G. (2002). The market of borrowing stocks. *Journal of Financial Economics*, 26:271–306.

[55] Davydov, G. and Linetsky, V. (2001). Pricing and hedging path-dependent options under the cev process. *Management Science*, 47(7):949–965.

[56] Dayanik, S. and Karatzas, I. (2003). On the optimal stopping problem for one-dimensional diffusions. *Stochastic Processes and Their Applications*, 107:173–212.

[57] Diebold, F. and Rudebusch, G. (2013). *Yield Curve Modeling and Forecasting: A Dynamic Nelson-Siegel Approach*. Princeton University Press.

[58] Diebold, F., Rudebusch, G., and Aruoba, B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics*, 131:309–338.

[59] Diether, K. B., Lee, K., and Werner, I. M. (2009). Short-sale strategies and return predictability. *Review of Financial Studies*, 22(2):575–607.

[60] Duffie, D. (2010). The failure mechanics of dealer banks. *Journal of Economic Perspectives*, 24(1):51–72.

[61] Duffie, D., Filipovic, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability*, 13(3):984–1053.

[62] Duffie, D., Garleanu, N., and Pedersen, L. H. (2002). Securities lending, shorting and pricing. *Journal of Financial Economics*, 66:307–339.

[63] Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406.

[64] Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376.

[65] Duffie, D. and Singleton, K. (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12:687–720.

[66] Dyrssen, H., Ekstrom, E., and Tysk, J. (2014). Pricing equations in jump-to-default models. *International Journal of Theoretical and Applied Finance*, 17(3).

[67] Eo, Y. and Kang, K. (2014). Forecasting the term structure of interest rates with potentially mis-specified models. Working Paper.

[68] Erakar, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *Journal of Finance*, 59(3):1367–1403.

[69] Fonseca, J. D., Grasselli, M., and Tebaldi, C. (2008). A multifactor volatility heston model. *Quantitative Financ*, 8(6):591–604.

[70] Fouque, J. and Lorig, M. (2011). A fast mean-reverting correction to heston's stochastic volatility model. *SIAM Journal on Financial Mathematics*, 2(1):221–254.

[71] Fouque, J., Papanicolaou, G., and Sircar, R. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press.

[72] Fouque, J., Papanicolaou, G., Sircar, R., and Solna, K. (2003). Multiscale stochastic volatility asymptotics. *Multiscale Modeling & Simulations*, 2(1):22–42.

[73] Fouque, J., Papanicolaou, G., Sircar, R., and Solna, K. (2004). Maturity cycles in implied volatility. *Finance and Stochastics*, 8(4):451–477.

[74] Fukasawa, M. (2011). Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15(4):635–654.

[75] Fukasawa, M., I, I., Maghrebi, N., Oya, K., Ubukata, M., and Yamazaki, K. (2011). Model-free implied volatility: from surface to index. *International Journal of Theoretical and Applied Finance*, 14(4):433–463.

[76] Gauthier, P. and Rivaille, P. (2009). Fitting the smile, smart parameters for sabr and heston. Working Paper.

[77] Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164:130–141.

[78] Geweke, J. and Amisano, G. (2012). Prediction with misspecified models. *American Economic Review*, 102(3):482–486.

[79] Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38:481–482.

[80] Glasserman, P. and Nouri, B. (2012). Contingent capital with a capital-ratio trigger. *Management Science*, 58:1816–1833.

[81] Guo, X. and Liu, J. (2005). Stopping at the maximum of geometric brownian motion when signals are received. *Journal of Applied Probability*, 42:826–838.

[82] Guo, X. and Zhang, Q. (2005). Optimal selling rules in a regime-switching model. *IEEE Transactions on Automatic Control*, 50:1450–1455.

[83] Gupta, A., Akuzawa, T., and Nishiyama, Y. (2013). Quantitative evaluation of contingent capital and its applications. *North American Journal of Economics and Finance*, 26:457–486.

[84] Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D. (2003). Managing smile risk. *Wilmott Magazine*, pages 84–108.

[85] Henry-Labordere, P. (2005). A general asymptotic implied volatility for stochastic volatility models. Working Paper.

[86] Heston, S. (1993). Closed-from solution for options with stochastic volatility with application to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

[87] Hoeting, J., Madigan, D., Raftey, A., and Volinsky, C. (1999). Bayesian model averaging. *Statistical Science*, 14:382–401.

[88] Jobert, A. and Rogers, L. (2006). Option pricing with markov-modulated dynamics. *SIAM Journal on Control and Optimization*, 44:2063–2078.

[89] Johannes, M. and Polson, N. (2009). *Particle Filtering, Handbook of Financial Time Series*. Springer-Verlag Berlin Heidelberg.

[90] Jones, C. M. and Lamont, O. A. (2002). Short-sale constraints and stock returns. *Journal of Financial Economics*, 66:207–239.

[91] Joslin, S., Le, A., and Singleton, K. (2013). Why gaussian macro-finance term structure models are (nearly) unconstrained factor-vars. *Journal of Financial Economics*, 109(3):604–622.

[92] Joslin, S., Priebsch, M., and Singleton, K. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, 69(3):1197–1233.

[93] Kallsen, J., Muhle-Karbe, J., and Vob, M. (2010). Pricing options on variance in affine stochastic volatility models. *Mathematical Finance*, 21(4):627–641.

[94] Keller-Ressel, M., Schachermayer, M., and Teichmann, J. (2011). Affine processes are regular. *Probability Theory and Related Fields*, 151:591–611.

[95] Kijima, M., Suzuki, T., and Tanaka, K. (2009a). A latent process model for the pricing of corporate securities. *Mathematical Method for Operation Research*, 69:439–455.

[96] Kijima, M., Tanaka, K., and Wong, T. (2009b). A multi-quality model of interest rates. *Quantitative Finance*, 9(2):133–145.

[97] Kim, D. and Singleton, K. (2012). Term structure models and the zero bound: an empirical investigation of japanese yields. *Journal of Econometrics*, 170:32–49.

[98] Lando, D. (1998). On cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120.

[99] Leippold, M. and Wu, L. (2002). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis*, 37(2):271–295.

[100] Li, G. and Zhang, C. (2010). On the number of state variables in option pricing. *Management Science*, 56(11):2058–2075.

[101] Li, H., Tao, L., and Yu, C. (2013). No-arbitrage taylor rules with switching regimes. *Management Science*, 59(10):2278–2294.

[102] Linz, P. (1969). Numerical methods for volterra integral equations of the first kind. *The Computer Journal*, 12(4):393–397.

[103] Longstaff, F. A. and Schwartz, E. S. (1995). A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance*, 50(3):789–819.

[104] Lord, R. and Kahl, C. (2007). Optimal fourier inversion in semi-analytical option pricing. *Journal of Computational Finance*, 10(4):1–30.

[105] Mendoza-Arriaga, R., Carr, P., and Linetsky, V. (2010). Time-changed markov processes in unified credit-equity modeling. *Mathematical Finance*, 20(4):527–569.

[106] Nagashima, K., Chung, T., and Tanaka, K. (2014). Asymptotic expansion formula of option price under multifactor heston model. *Asia-Pacific Financial Markets*, 21:351–396.

[107] Negro, M. D., Hasegawa, R., and Schorfheide, F. (2014). Dynamic prediction pools: an ingestigation of financial frictions and forecasting performance. NBER Working Paper No. 20575.

[108] Negro, M. D. and Schorfheide, F. (2010). *Bayesian Macroeconometrics*. Handbook of Bayesian Econometrics.

[109] Nualart, D. (2006). *The Malliavin calculus and related topics*. Springer.

[110] Osajima, Y. (2007). The asymptotic expansion formula of implied volatility for dynamic sabr model and fx hybrid model. Working Paper.

[111] Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63:3–50.

[112] Papageorgiou, E. and Sircar, R. (2009). Multiscale intensity models and name grouping for valuation of multi-name credit derivatives. *Applied Mathematical Finance*, 16(4):353–383.

[113] Pennacchi, G. (2011). A structural model of contingent bank capital. Federal Reserve Bank of Cleveland Working Paper 10-04.

[114] Peskir, G. and Shiryaev, A. (2005). *Optimal Stopping and Free-Boundary Problems*. Birkhauser Verlag.

[115] Piazzesi, M. (2010). *Affine Term Structure Models, Handbook of Financial Econometrics*. North Holland, Elsevier.

[116] Pliska, S. and Ye, J. (2007). Optimal life insurance purchase and consumption/investment under uncertain lifetime. *Journal of Banking and Finance*, 31:1307–1319.

[117] Rogers, L. and Williams, D. (2000). *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations*. Cambridge University Press.

[118] Spiegeleer, J. D. and Schoutens, W. (2012). Pricing contingent convertibles: A derivatives approach. *Journal of Derivatives*, 20(2):27–36.

[119] Takahashi, A. and Yamada, T. (2012). An asymptotic expansion with push-down of malliavin weights. *SIAM Journal on Financial Mathematics*, 3(1):95–136.

[120] Tanaka, K. (2012). Irreversible investment with regime switching: Revisit with linera algebra. Working Paper, Tokyo Metropolitan University.

[121] Tanaka, K., Yamada, T., and Watanabe, T. (2010). Applications of gram-charlier expansion and bond moments for pricing of interest rates and credit risk. *Quantitative Finance*, 10(6):645–662.

[122] Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

[123] Waggoner, D. and Zha, T. (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics*, 171:167–184.

[124] Wright, J. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review*, 101:1514–1534.

[125] Zhang, B., Zhou, H., and Zhu, H. (2009). Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *Review of Financial Studies*, 22(12):5099–5131.

[126] Zheng, W. and Kwok, Y. (2014). Closed form pricing formulas for discretely sampled generalized variance swaps. *Mathematical Finance*, 24(4):855–881.