ESSAYS IN FINANCIAL AND INSURANCE MATHEMATICS

by

Xueying Hu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Erhan Bayraktar, Co-Chair
Professor Haitao Li, Co-Chair
Professor Joseph G. Conlon
Professor Virginia R. Young
Associate Professor Kristen S. Moore

To my beloved parents: Changhong Sun and Jingwen Hu.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the enormous support and encouragement from my Ph.D. advisors Professor Erhan Bayraktar and Professor Haitao Li. Not only did they teach me precious knowledge on mathematical and quantitative finance, but also influenced me with their passion toward research, which became a life-long asset of mine. I thank Professor Bayraktar for providing many opportunities on my professional growth, for his patient and support during difficult times, for being a role model on positive attitude and time management, and for many things and beyond. I thank Professor Li for his helpful career advice, for providing me with research freedom, and for his guidance when necessary. I am very grateful for all your insights and advice during the past five years.

I am deeply grateful for my committee members for their great help throughout my Ph.D studies. I am grateful for professor Joseph Conlon's stochastic process and computational finance classes and for his inspirational reading course on Stochastic Control Theory. I also would like to express my appreciation to professor Virginia R. Young for her careful line-by-line revisions and helpful discussions on my first journal paper which became the first part of this thesis. In addition, I thank professor Kristen Moore for her help on the Markov Chain Approximation Method.

Furthermore, I want to thank Professors Kay Giesecke at Stanford University and Professor Tao Li at the City University of Hong Kong for the beneficial discussions and their insightful comments on my subsequent two papers listed in the second and

third part of my thesis respectively.

I feel honored for having been part of the Department of Mathematics at the University of Michigan. I thank you for providing me with an academic home and proper financial support during the last five years. Here I want to thank Professors Martias Jonsson, Semih Sezer, and Jinho Baik for having me as their Teaching Assistant. I also thank all the participants at the AIM and Financial Math seminars for providing me with opportunities to talk about my research and to learn from their talks. I also thank my colleagues, Hao Xing, Ting Wang, Bo Yang, Lu Lu, Zhengjie Xu, Song Yao, and Lei Wang for sharing their graduate study experience with me and for discussing many aspects in life with me as international students.

Special thanks to my friends, Lina Xiong, Chunhui Yu, Jie Zeng, Jianwei Li, Youna Hu, Ting Lu, Zhanyang Zhang, Yang Zhang, Jing Zhou, Yijie Zhou, Fanhuan Zhou, Zachary Scherr, Tomoki Ohsawa, Henry Boateng, Holly Chung, Dawei Shen, Hao Chen, and many others who have have been by my side encouraging me during difficult times, and who have also been sharing many joyfulness in life in the past five years. Pursuing my Ph.D. degree would have been rather lonely without your company.

I cannot finish without saying how grateful I am for my family's support: my parents Jingwen Hu and Changhong Sun, aunt Kai Tan, uncle Changzheng Sun, and cousin Lihe Sun. Their unconditional love and constant support have made me who I am today.

To them I dedicate this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

This thesis consists of three essays.

In the first essay, chapter II, we investigate the optimal investment strategy for an individual to minimize the probability of going bankruptcy before death. The individual is assumed to invest in a financial market with one riskfree and one risky asset, with the latter's price following a diffusion with stochastic volatility. Using stochastic optimal control techniques, perturbation analysis and Markov Chain Approximation method, we solve this minimization problem and provide insights for retirees.

This chapter is based on Bayraktar et al. [16]. Parts of this work as been presented at SIAM Conference on Financial Mathematics and Engineering (FM10), San Francisco, November 19, 2010; First Annual University of Michigan SIAM Student Conference, Ann Arbor, November 13, 2010; Bachelier Finance Society World Congress, Fields Institute, Toronto, June 25, 2010; and Financial/Actuarial Mathematics Seminar, University of Michigan, Ann Arbor, 2009.

Chapter III extends the Heston stochastic volatility model to include state-dependent jumps in the price and the volatility, and develops a method for the exact simulation of this model. The jumps arrive with a stochastic intensity that may depend on

time, price, volatility and jump counts. They may have an impact on the price or the volatility, or both. The random jump size may depend on the price and volatility. Numerical experiments demonstrate the performance of our exact method.

This chapter is based on Bayraktar et al. [15]. Parts of this work has been presented at Statistic Student Seminar, University of Michigan, Ann Arbor, October 13, 2011 and will be presented at SIAM Conference on Financial Mathematics and Engineering (FM12), Minnesota, July 2012.

In Chapter IV, we study the properties of systemic sovereign credit risk using CDS spreads for U.S. and major sovereign countries. We develop a regime-switching two-factor model that allows for both global-systemic and sovereign-specific credit shocks, and use maximum likelihood estimation to calibrate model parameters to CDS weekly data. The preliminary results suggest a heterogeneity across different countries in their sensitivity to system risk. In addition, the regime across high-volatility and low-volatility regimes behaves differently with asymmetric regime-shift probabilities. This chapter is based on Li et al. [72].

CHAPTER II

Minimizing the Probability of Lifetime Ruin under Stochastic Volatility

2.1 Introduction

Pension actuaries traditionally have computed the liabilities for defined benefit (DB) pension plans; however, more and more employees are participating in defined contribution (DC) plans. Indeed, in June 2007, the Employee Benefits Research Institute (EBRI) reported that in 1979, among active workers participating in retirement plans, the percentages in DB plans only, DC plans only, and both DB and DC plans were 62%, 16%, and 22%, respectively. The corresponding percentages in 2005 were 10%, 63%, and 27%, respectively.

In terms of numbers of employees, EBRI reported that in 1980, 30.1 million active workers participated in DB plans, while 18.9 million workers participated in DC plans. The corresponding numbers in 2004 were 20.6 and 52.2 active workers, respectively. Finally, in terms of numbers of plans in the private sector, in 1980, there were 148 thousand DB plans and 341 thousand DC plans; the corresponding numbers in 2004 were 47 and 653 thousand plans, respectively.

Therefore, however one measures the change in employee coverage under DB versus DC plans, it is clear that pension actuaries will need to adapt to the migration from DB to DC plans. One way that they can adapt is to switch from advising

employers about their DB liabilities to providing investment advice for retirees and employees in DC plans. The purpose of our proposed research is to help train actuaries for this opportunity under the easy-to-explain goal of an employee or retiree avoiding bankruptcy.

Previous work focused on finding the optimal investment strategy to minimize the probability of bankruptcy under a variety of situations: (1) allowing the individual to invest in a standard Black-Scholes financial market with a rate of consumption given by some function of wealth, Young [97], Bayraktar and Young [19]; (2) incorporating immediate and deferred annuities in the financial market, Milevsky et al. [81], Bayraktar and Young [22]; (3) limiting borrowing or requiring that borrowing occur at a higher rate than lending, Bayraktar and Young [20]; (4) modeling consumption as an increasing function of wealth or as a random process correlated with the price process of the stock, Bayraktar and Young [21], Bayraktar et al. [17], Bayraktar and Young [18]. Throughout this body of work, the price process of the stock is modeled as a geometric Brownian motion, which is arguably unrealistic, but has given results that one can consider to be "first approximations." Here we extend some of the previous work and allow the stock price to exhibit stochastic volatility. Additionally, we intend to find easy-to-implement rules that will result in nearly minimal probabilities of bankruptcy under stochastic volatility.

The rest of the chapter is organized as follows. In Section 2, we introduce the financial market and define the problem of minimizing the probability of lifetime ruin. In Section 3, we present a related optimal controller-stopper problem, and show that the solution of that problem is the Legendre dual of the minimum probability of lifetime ruin. By solving the optimal controller-stopper problem, we effectively solve the problem of minimizing the probability of lifetime ruin. Relying on the results in

Section 3, we find an asymptotic approximation of the minimum probability of ruin and the optimal strategy in Section 4. On the other hand, in Section 5, relying on the Markov Chain Approximation Method, we construct a numerical algorithm that solves the original optimal control problem numerically. In Section 6, we present some numerical experiments.

We learn that the optimal investment strategy in the presence of stochastic volatility is not necessarily to invest less in the risky asset than when volatility is fixed. We also observe that the minimal probability of ruin can be almost attained by the asymptotic approximation described in Section 5.1. Also, if an individual uses the investment prescribed by the optimal investment strategy for the constant volatility environment while updating the volatility variable in this formula according to her observations, it turns out she can almost achieve the minimum probability of ruin in a stochastic volatility environment.

2.2 The Financial Market and the Probability of Lifetime Ruin

In this section, we present the financial ingredients that make up the individual's wealth, namely, consumption, a riskless asset, and a risky asset. We, then, define the minimum probability of lifetime ruin.

We assume that the individual invests in a riskless asset whose price at time t, X_t, follows the process $dX_t = rX_t dt, X_0 = x > 0$, for some fixed rate of interest $r > 0$. Also, the individual invests in a risky asset whose price at time t, S_t, follows a diffusion given by

$$\mathrm{d}S_t = S_t \left(\mu dt + \sigma_t \, \mathrm{d}B_t^{(1)} \right), \quad S_0 = S > 0, \tag{2.2.1}$$

in which $\mu > r$ and σ_t is the (random) volatility of the price process at time t. Here, $B^{(1)}$ is a standard Brownian motion with respect to a filtered probability space

$(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F} = \{\mathcal{F}_t\}_{t\geq 0})$. We assume that the stochastic volatility is given by

$$\sigma_t = f(Y_t, Z_t), \tag{2.2.2}$$

in which f is a smooth positive function that is bounded and bounded away from zero, and Y and Z are two diffusions. Below, we follow Fouque et al. [44] in specifying the dynamics of Y and Z. Note that if f is constant, then S follows geometric Brownian motion, and that case is considered by Young [97].

The first diffusion Y is a fast mean-reverting Gaussian Ornstein-Uhlenbeck process. Denote by $1/\epsilon$ the rate of mean reversion of this process, with $0 < \epsilon \ll 1$ corresponding to the time scale of the process. Y is an ergodic process, and we assume that its invariant distribution is independent of ϵ. In particular, the invariant distribution is normal with mean m and variance ν^2. The resulting dynamics of Y are given by

$$\mathrm{d}Y_t = \frac{1}{\epsilon}\,(m - Y_t)\,dt + \nu\sqrt{\frac{2}{\epsilon}}\,\mathrm{d}B_t^{(2)}, \quad Y_0 = y \in \mathbf{R}, \tag{2.2.3}$$

in which $B^{(2)}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F})$. Suppose $B^{(1)}$ and $B^{(2)}$ are correlated with (constant) coefficient $\rho_{12} \in (-1, 1)$.

Under its invariant distribution $\mathcal{N}(m, \nu^2)$, the autocorrelation of Y is given by

$$\mathbb{E}\left[(Y_t - m)(Y_s - m)\right] = \nu^2\, e^{-|t-s|/\epsilon}. \tag{2.2.4}$$

Therefore, the process decorrelates exponentially fast on the time scale ϵ; thus, we refer to Y as the fast volatility factor.

The second factor Z driving the volatility of the risky asset's price process is a slowly varying diffusion process. We obtain this diffusion by applying the time change $t \to \delta \cdot t$ to a given diffusion process:

$$\mathrm{d}\widetilde{Z}_t = g(\widetilde{Z}_t)\,dt + h(\widetilde{Z}_t)\,\mathrm{d}\tilde{B}_t, \tag{2.2.5}$$

in which $0 < \delta \ll 1$ and $\tilde{B}$ is a standard Brownian motion. The coefficients g and h are smooth and at most linearly growing at infinity, so (2.2.5) has a unique strong solution. Under the time change $t \to \delta \cdot t$, define $Z_t = \widetilde{Z}_{\delta \cdot t}$. Then, the dynamics of Z are given by

$$\mathrm{d}Z_t = \delta\, g(Z_t)\, dt + h(Z_t)\, \mathrm{d}\tilde{B}_{\delta \cdot t}, \quad Z_0 = z \in \mathbf{R}. \tag{2.2.6}$$

In distribution, we can write these dynamics as

$$\mathrm{d}Z_t = \delta\, g(Z_t)\, dt + \sqrt{\delta}\, h(Z_t)\, \mathrm{d}B_t^{(3)}, \quad Z_0 = z \in \mathbf{R}, \tag{2.2.7}$$

in which $B^{(3)}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F})$. Suppose $B^{(1)}$ and $B^{(3)}$ are correlated with (constant) coefficient $\rho_{13} \in [-1, 1]$. Similarly, suppose $B^{(2)}$ and $B^{(3)}$ are correlated with (constant) coefficient $\rho_{23} \in [-1, 1]$. To ensure that the covariance matrix of the Brownian motions is positive semi-definite, we impose the following condition on the ρ's:

$$1 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 \geq 0. \tag{2.2.8}$$

Let W_t be the wealth at time t of the individual, and let π_t be the amount that the decision maker invests in the risky asset at that time. It follows that the amount invested in the riskless asset is $W_t - \pi_t$. We assume that the individual consumes at a constant rate $c > 0$. Therefore, the wealth process follows

$$\mathrm{d}W_t = [rW_t + (\mu - r)\pi_t - c]\, dt + f(Y_t, Z_t)\, \pi_t\, \mathrm{d}B_t^{(1)}, \tag{2.2.9}$$

and we suppose that initial wealth is non-negative; that is, $W_0 = w \geq 0$.

By *lifetime ruin*, we mean that the individual's wealth reaches zero before she dies. Define the corresponding hitting time by $\tau_0 := \inf\{t \geq 0 : W_t \leq 0\}$. Let τ_d denote the random time of death of the individual, which is independent of the Brownian motions. We assume that τ_d is exponentially distributed with parameter

λ (that is, with expected time of death equal to $1/\lambda$); this parameter is also known as the *hazard rate*, or, *force of mortality*.

Moore and Young [82] minimize the probability of lifetime ruin with varying hazard rate and show that by updating the hazard rate each year and treating it as a constant, the individual can quite closely obtain the minimal probability of ruin when the true hazard rate is Gompertz. Specifically, at the beginning of each year, set λ equal to the inverse of the individual's life expectancy at that time. Compute the corresponding optimal investment strategy as given below, and apply that strategy for the year. According to the work of Moore and Young [82], this scheme results in a probability of ruin close to the minimum probability of ruin. Therefore, there is no significant loss of generality to assume that the hazard rate is constant and revise its estimate each year.

Denote the minimum probability of lifetime ruin by $\psi(w, y, z)$, in which the arguments w, y, and z indicate that one conditions on the individual possessing wealth w at the current time with the two factors Y and Z taking the values y and z, respectively, then. Thus, ψ is the minimum probability that $\tau_0 < \tau_d$, in which one minimizes with respect to admissible investment strategies π. A strategy π is *admissible* if it is $\mathcal{F}_t$-progressively measurable, and if it satisfies the integrability condition $\int_0^t \pi_s^2 \, ds < \infty$ almost surely for all $t \geq 0$. Thus, ψ is formally defined by

$$\psi(w, y, z) = \inf_{\pi} \mathbb{P}^{w,y,z} \left[\tau_0 < \tau_d \right]. \tag{2.2.10}$$

Here, $\mathbb{P}^{w,y,z}$ indicates the probability conditional on $W_0 = w$, $Y_0 = y$, and $Z_0 = z$. Note that if $w \geq c/r$, then $\psi(w, y, z) = 0$ because the individual can invest c/r of her wealth in the riskless asset and generate a rate of income equal to c, which exactly covers her consumption. Therefore, we effectively only need to determine the minimum probability of lifetime ruin and corresponding optimal investment strategy

on the domain $\mathbf{D} := \{(w, y, z) \in \mathbf{R}^3 : w \in [0, c/r]\}$.

2.3 Computing the Minimum Probability of Lifetime Ruin

2.3.1 A related optimal controller-stopper problem

In this section, we present an optimal controller-stopper problem whose solution $\hat{\psi}$ is the Legendre dual of the minimum probability of ruin ψ. It is not clear *a priori* that the value function ψ is convex or smooth due to the implicit dependent on the initial values of the state variable. By passing to the controller-stopper problem, however, we can obtain the regularity of $\hat{\psi}$ in more simply, which, in turn, provides an intermediate tool in the proof of regularity of ψ. The dual relationship and the analysis of the controller-stopper problem are, therefore, crucial and worth investigating.

First, note that we can represent the three Brownian motions from Section 2 as follows: given $B^{(1)}$, $B^{(2)}$, and $B^{(3)}$, define $\widetilde{B}^{(1)}$, $\widetilde{B}^{(2)}$, and $\widetilde{B}^{(3)}$ via the following invertible system of equations:

$$\begin{aligned}
B_t^{(1)} &= \widetilde{B}_t^{(1)}, \\
B_t^{(2)} &= \rho_{12}\, \widetilde{B}_t^{(1)} + \sqrt{1-\rho_{12}^2}\, \widetilde{B}_t^{(2)}, \\
B_t^{(3)} &= \rho_{13}\, \widetilde{B}_t^{(1)} + \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}}\, \widetilde{B}_t^{(2)} + \frac{\sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2) - (\rho_{23} - \rho_{12}\rho_{13})^2}}{\sqrt{1-\rho_{12}^2}}\, \widetilde{B}_t^{(3)}.
\end{aligned} \tag{2.3.1}$$

One can show that $\widetilde{B}^{(1)}$, $\widetilde{B}^{(2)}$, and $\widetilde{B}^{(3)}$ thus defined are three *independent* standard Brownian motions on $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F})$. Also notice that condition (2.2.8) on the ρ's guarantees that the expression under the square root in the coefficient of $\widetilde{B}_t^{(3)}$ is non-negative.

Next, define the controlled process X^γ by

$$\mathrm{d}X_t^\gamma = -(r-\lambda)\,X_t^\gamma\,dt - \frac{\mu - r}{f(Y_t, Z_t)}\,X_t^\gamma\,\mathrm{d}\widetilde{B}_t^{(1)} + \gamma_t^{(2)}\mathrm{d}\widetilde{B}_t^{(2)} + \gamma_t^{(3)}\mathrm{d}\widetilde{B}_t^{(3)}, \quad X_0 = x > 0, \tag{2.3.2}$$

in which $\gamma = \left(\gamma^{(2)}, \gamma^{(3)}\right)$ is the control, and Y and Z are given in (2.2.3) and (2.2.7), respectively.

For $x > 0$, define the function $\hat{\psi}$ by

$$\hat{\psi}(x, y, z) = \inf_{\tau}\sup_{\gamma}\mathbb{E}^{x,y,z}\left[\int_0^\tau e^{-\lambda t} c\,X_t^\gamma\,dt + e^{-\lambda\tau}\min\left((c/r)X_\tau^\gamma, 1\right)\right]. \tag{2.3.3}$$

$\hat{\psi}$ is the value function for an optimal controller-stopper problem. Indeed, the controller chooses among processes γ in order to maximize the discounted running "penalty" to the stopper given by $c\,X_t^\gamma$ in (2.3.3). On the other hand, the stopper chooses the time to stop the game in order to minimize the penalty but has to incur the terminal cost of $\min\left((c/r)X_\tau^\gamma, 1\right)$, discounted by $e^{-\lambda\tau}$ when she stops.

Bayraktar and Young [18] consider a controller-stopper problem that is mathematically similar to the one in this chapter; see that paper for details of the following assertions–specifically, see Theorem 2.4 and its proof. One can show that the controller-stopper problem has a continuation region given by $\{(x, y, z) : 0 \leq x_{c/r}(y, z) \leq x \leq x_0(y, z)\}$ for some functions $0 \leq x_{c/r}(y, z) \leq r/c \leq x_0(y, z)$ with $(y, z) \in \mathbf{R}^2$. Thus, if $x \leq x_{c/r}(y, z)$, we have $\hat{\psi}(x, y, z) = (c/r)\,x$, and if $x \geq x_0(y, z)$, we have $\hat{\psi}(x, y, z) = 1$. Moreover, $\hat{\psi}$ is non-decreasing and concave with respect to x on $\mathbf{R}^+$ (increasing and strictly concave in the continuation region) and is the unique classical solution of the following free-boundary problem on $\left[x_{c/r}(y, z),\, x_0(y, z)\right]$:

$$\begin{cases} cx + \left(\frac{1}{\epsilon}\,\mathcal{L}_0 + \frac{1}{\sqrt{\epsilon}}\,\mathcal{L}_1 + \mathcal{L}_2 + \sqrt{\delta}\,\mathcal{M}_1 + \delta\,\mathcal{M}_2 + \sqrt{\frac{\delta}{\epsilon}}\,\mathcal{M}_3\right)\hat{\psi} + NL^{\epsilon,\delta} = 0; \\ \hat{\psi}(x_{c/r}(y, z), y, z) = \frac{c}{r}\,x_{c/r}(y, z), \quad \hat{\psi}_x(x_{c/r}(y, z), y, z) = \frac{c}{r}; \\ \hat{\psi}(x_0(y, z), y, z) = 1, \quad \hat{\psi}_x(x_0(y, z), y, z) = 0; \end{cases} \tag{2.3.4}$$

in which

$$\mathcal{L}_0 v = (m-y)\, v_y + \nu^2\, v_{yy}, \tag{2.3.5}$$

$$\mathcal{L}_1 v = -\rho_{12}\, \frac{\mu - r}{f(y,z)}\, \nu\, \sqrt{2}\, x\, v_{xy}, \tag{2.3.6}$$

$$\mathcal{L}_2 v = -\lambda\, v - (r-\lambda)\, x\, v_x + \frac{1}{2} \left(\frac{\mu - r}{f(y,z)} \right)^2 x^2\, v_{xx}, \tag{2.3.7}$$

$$\mathcal{M}_1 v = -\rho_{13}\, \frac{\mu - r}{f(y,z)}\, h(z)\, x\, v_{xz}, \tag{2.3.8}$$

$$\mathcal{M}_2 v = g(z)\, v_z + \frac{1}{2}\, h^2(z)\, v_{zz}, \tag{2.3.9}$$

$$\mathcal{M}_3 v = \rho_{23}\, \nu\, \sqrt{2}\, h(z)\, v_{yz}, \tag{2.3.10}$$

and

$$\begin{aligned} NL^{\epsilon,\delta} = \sup_{\gamma} \Bigg[& \frac{1}{2} \left(\left(\gamma^{(2)}\right)^2 + \left(\gamma^{(3)}\right)^2 \right) \hat{\psi}_{xx} \\ & + \gamma^{(2)} \left(\nu \sqrt{\frac{2}{\epsilon}}\, \sqrt{1-\rho_{12}^2}\, \hat{\psi}_{xy} + \sqrt{\delta}\, h(z)\, \frac{\rho_{23} - \rho_{12}\rho_{23}}{\sqrt{1-\rho_{12}^2}}\, \hat{\psi}_{xz} \right) \\ & + \gamma^{(3)}\, \sqrt{\delta}\, h(z)\, \frac{\sqrt{(1-\rho_{12}^2)(1-\rho_{13}^2) - (\rho_{23} - \rho_{12}\rho_{13})^2}}{\sqrt{1-\rho_{12}^2}}\, \hat{\psi}_{xz} \Bigg]. \end{aligned} \tag{2.3.11}$$

Because $\hat{\psi}$ is concave with respect to x, we can express $NL^{\epsilon,\delta}$ as follows:

$$NL^{\epsilon,\delta} = -\frac{1}{\epsilon}\, \nu^2 \left(1-\rho_{12}^2\right) \frac{\hat{\psi}_{xy}^2}{\hat{\psi}_{xx}} - \frac{1}{2}\, \delta\, h^2(z) \left(1-\rho_{13}^2\right) \frac{\hat{\psi}_{xz}^2}{\hat{\psi}_{xx}} - \nu\sqrt{2}\, \sqrt{\frac{\delta}{\epsilon}}\, h(z) \left(\rho_{23} - \rho_{12}\rho_{13}\right) \frac{\hat{\psi}_{xy}\hat{\psi}_{xz}}{\hat{\psi}_{xx}}. \tag{2.3.12}$$

2.3.2 Convex Legendre dual of $\hat{\psi}$

Since $\hat{\psi}$ is strictly concave with respect to x in its continuation region (which corresponds to wealth lying in $[0, c/r]$), we can define its convex dual Ψ by the Legendre transform: for $(w,y,z) \in \mathbf{D} = \{(w,y,z) \in \mathbf{R}^3 : w \in [0, c/r]\}$,

$$\Psi(w,y,z) = \max_x \left(\hat{\psi}(x,y,z) - wx \right). \tag{2.3.13}$$

In this section, we show that the convex dual Ψ is the minimum probability of lifetime ruin; then, in the next section, we asymptotically expand $\hat{\psi}$ in powers of $\sqrt{\epsilon}$ and $\sqrt{\delta}$.

Theorem 2.3.1. *Ψ equals the minimum probability of lifetime ruin ψ on $\mathbf{D}$, and the investment policy π^* given in feedback form by $\pi_t^* = \pi^*(W_t^*, Y_t, Z_t)$ is an optimal policy, in which W^* is the optimally controlled wealth (that is, wealth controlled by π^*) and the function π^* is given by*

$$\pi^*(w,y,z) = -\frac{\mu - r}{f^2(y,z)}\,\frac{\psi_w}{\psi_{ww}} - \rho_{12}\sqrt{\frac{2}{\epsilon}}\,\frac{\nu}{f(y,z)}\,\frac{\psi_{wy}}{\psi_{ww}} - \rho_{13}\sqrt{\delta}\,\frac{h(z)}{f(y,z)}\,\frac{\psi_{wz}}{\psi_{ww}}, \quad (2.3.14)$$

in which the right-hand side of (2.3.14) is evaluated at (w,y,z).

Proof. From (2.3.13), it follows that the critical value x^* solves $w = \hat{\psi}_x(x,y,z)$; thus, given w, we have $x^* = I(w,y,z)$, in which I is the inverse function of $\hat{\psi}_x$ with respect to x. Therefore, $\Psi(w,y,z) = \hat{\psi}(I(w,y,z),y,z) - wI(w,y,z)$. By differentiating this expression of Ψ with respect to w, we obtain $\Psi_w(w,y,z) = \hat{\psi}_x(I(w,y,z),y,z)I_w(w,y,z) - I(w,y,z) - wI_w(w,y,z) = -I(w,y,z)$; thus, $x^* = -\Psi_w(w,y,z)$. Similarly, we obtain (with $w = \hat{\psi}_x(x,y,z)$) the following expressions:

$$\hat{\psi}_{xx}(x,y,z) = -\frac{1}{\Psi_{ww}(w,y,z)}, \quad (2.3.15)$$

$$\hat{\psi}_y(x,y,z) = \Psi_y(w,y,z), \quad (2.3.16)$$

$$\hat{\psi}_z(x,y,z) = \Psi_z(w,y,z), \quad (2.3.17)$$

$$\hat{\psi}_{yy}(x,y,z) = \Psi_{wy}(w,y,z)\,\hat{\psi}_{xy}(x,y,z) + \Psi_{yy}(w,y,z), \quad (2.3.18)$$

$$\hat{\psi}_{zz}(x,y,z) = \Psi_{wz}(w,y,z)\,\hat{\psi}_{xz}(x,y,z) + \Psi_{zz}(w,y,z), \quad (2.3.19)$$

$$\hat{\psi}_{xy}(x,y,z) = \Psi_{wy}(w,y,z)\,\hat{\psi}_{xx}(x,y,z), \quad (2.3.20)$$

$$\hat{\psi}_{xz}(x,y,z) = \Psi_{wz}(w,y,z)\,\hat{\psi}_{xx}(x,y,z), \quad (2.3.21)$$

and

$$\hat{\psi}_{yz}(x,y,z) = \Psi_{wy}(w,y,z)\,\Psi_{wz}(w,y,z)\,\hat{\psi}_{xx}(x,y,z) + \Psi_{yz}(w,y,z). \quad (2.3.22)$$

By substituting $x^* = -\Psi_w(w,y,z)$ into the free-boundary problem for $\hat{\psi}$, namely (2.3.4), one can show that Ψ uniquely solves the following boundary-value problem on $\mathbf{D}$:

$$\begin{cases} \min_\beta \mathcal{D}^\beta v(w,y,z) = 0; \\ \\ v(0,y,z) = 1, \quad v(c/r,y,z) = 0. \end{cases} \tag{2.3.23}$$

where the differential operator $\mathcal{D}^\beta$ is given by

$$\begin{aligned} \mathcal{D}^\beta v = &-\lambda\, v + (rw + (\mu - r)\beta - c)\, v_w + \frac{1}{\epsilon}\,(m-y)\, v_y + \delta\, g(z)\, v_z \\ &+ \frac{1}{2} f^2(y,z)\,\beta^2\, v_{ww} + \frac{1}{\epsilon}\,\nu^2\, v_{yy} + \frac{1}{2}\,\delta\, h^2(z)\, v_{zz} + \rho_{12}\, f(y,z)\,\beta\,\nu\,\sqrt{\frac{2}{\epsilon}}\, v_{wy} \\ &+ \rho_{13}\, f(y,z)\,\beta\,\sqrt{\delta}\, h(z)\, v_{wz} + \rho_{23}\,\sqrt{2}\,\nu\,\sqrt{\frac{\delta}{\epsilon}}\, h(z)\, v_{yz}. \end{aligned} \tag{2.3.24}$$

Observe that Ψ is strictly convex in w because $\hat{\psi}$ is strictly concave in x in its continuation region which corresponds to $\mathbf{D}$ in the original space. Since Ψ is strictly convex with respect to w, the optimal policy π^* in (2.3.23) is given by the first-order necessary condition, which results in the expression in (2.3.14). Now, using a standard verification theorem we deduce that Ψ is the minimum probability of lifetime ruin ψ. □

Theorem 2.3.1 demonstrates the strong connection between $\hat{\psi}$ and ψ, namely that they are dual via the Legendre transform. (As an aside, if we have ψ, we can obtain $\hat{\psi}$ via $\hat{\psi}(x,y,z) = \min_w (\psi(w,y,z) + wx)$.) Therefore, if we have $\hat{\psi}$, then we obtain the minimum probability of ruin ψ via (2.3.13). More importantly, we get the optimal investment strategy π^* via (2.3.14). As a corollary to Theorem 2.3.1, we have the following expression for π^* in terms of the dual variable x.

Corollary 2.3.2. *In terms of the dual variable x, the optimal investment strategy π^* is given by $\pi_t^* = \hat{\pi}^*(X_t^*, Y_t, Z_t)$, in which X^* is the optimally controlled process*

X, and

$$\hat{\pi}^*(x,y,z) = -\frac{\mu - r}{f^2(y,z)}\, x\, \hat{\psi}_{xx} + \rho_{12}\sqrt{\frac{2}{\epsilon}}\,\frac{\nu}{f(y,z)}\,\hat{\psi}_{xy} + \rho_{13}\sqrt{\delta}\,\frac{h(z)}{f(y,z)}\,\hat{\psi}_{xz}, \quad (2.3.25)$$

with the right-hand side of (2.3.25) *evaluated at* (x,y,z).

Proof. Let $w = \hat{\psi}_x(x,y,z)$ in (2.3.14) and simplify the right-hand side via equations (2.3.15)-(2.3.22) to obtain (2.3.25). □

2.4 Asymptotic Approximation of the Minimum Probability of Lifetime Ruin

In this section, we asymptotically expand $\hat{\psi}$, the Legendre transform of the minimum probability of ruin, in powers of $\sqrt{\epsilon}$ and $\sqrt{\delta}$. (A parallel analysis of expanding the Legendre transform of the value function of the utility maximization problem was carried out in Jonsson and Sircar [64].) We expand $\hat{\psi}$ instead of ψ because if one were to do the latter, then one would note that each term in the expansion solves a *non-linear* differential equation. The differential equation for the zeroth-order term has a closed-form solution; however, none of the differential equations for the higher-order terms does. What this fact implies is that to solve any of these non-linear differential equations, one would have to assume that it has a convex solution, determine the corresponding linear free-boundary problem for the concave dual, solve this free-boundary problem, then invert the solution numerically, as in equation (2.3.13). Note that one would have to perform this procedure for *each* higher-order term in the expansion.

By contrast, when we expand $\hat{\psi}$, each term solves a *linear* differential equation, as we show below. We explicitly solve these linear differential equations, then invert the approximation using (2.3.13) *once* to obtain an approximation for the minimum probability of lifetime ruin ψ. Note that the resulting approximation of ψ is not

guaranteed to be a probability, that is, to lie in the interval $[0,1]$; however, our numerical experiments show that this is not a problem for the values of the parameters we consider. See [45] for an example of approximating a probability that solves a linear differential equation.

To begin, expand $\hat{\psi}$ and the free boundaries in powers of $\sqrt{\delta}$:

$$\hat{\psi} = \hat{\psi}_0 + \sqrt{\delta}\,\hat{\psi}_1 + \delta\,\hat{\psi}_2 + \cdots\,, \tag{2.4.1}$$

$$x_{c/r}(y,z) = x_{c/r,0}(y,z) + \sqrt{\delta}\,x_{c/r,1}(y,z) + \delta\,x_{c/r,2}(y,z) + \cdots\,, \tag{2.4.2}$$

and

$$x_0(y,z) = x_{0,0}(y,z) + \sqrt{\delta}\,x_{0,1}(y,z) + \delta\,x_{0,2}(y,z) + \cdots\,. \tag{2.4.3}$$

Insert the expression in (2.4.1) into $NL^{\epsilon,\delta}$ in (4.12) to obtain the following expansion in powers of $\sqrt{\delta}$:

$$\begin{aligned} NL^{\epsilon,\delta} = &-\frac{1}{\epsilon}\,\nu^2\left(1-\rho_{12}^2\right)\frac{\hat{\psi}_{0,xy}^2}{\hat{\psi}_{0,xx}} \\ &+\sqrt{\delta}\left[\frac{1}{\epsilon}\,\nu^2\left(1-\rho_{12}^2\right)\left(\left(\frac{\hat{\psi}_{0,xy}}{\hat{\psi}_{0,xx}}\right)^2\hat{\psi}_{1,xx} - 2\,\frac{\hat{\psi}_{0,xy}}{\hat{\psi}_{0,xx}}\,\hat{\psi}_{1,xy}\right)\right. \\ &\left.-\sqrt{\frac{2}{\epsilon}}\,\nu\,h(z)(\rho_{23}-\rho_{12}\rho_{13})\frac{\hat{\psi}_{0,xy}\,\hat{\psi}_{0,xz}}{\hat{\psi}_{0,xx}}\right] + \mathcal{O}(\delta). \end{aligned} \tag{2.4.4}$$

Keeping terms up to $\sqrt{\delta}$, we expand the free-boundary conditions in (2.3.4) as

$$\begin{aligned} &\hat{\psi}_0(x_{c/r,0}(y,z),y,z) + \sqrt{\delta}\left[x_{c/r,1}(y,z)\,\hat{\psi}_{0,x}(x_{c/r,0}(y,z),y,z) + \hat{\psi}_1(x_{c/r,0}(y,z),y,z)\right] \\ &\quad = \frac{c}{r}\left(x_{c/r,0}(y,z) + \sqrt{\delta}\,x_{c/r,1}(y,z)\right), \end{aligned} \tag{2.4.5}$$

$$\begin{aligned} &\hat{\psi}_{0,x}(x_{c/r,0}(y,z),y,z) + \sqrt{\delta}\left[x_{c/r,1}(y,z)\,\hat{\psi}_{0,xx}(x_{c/r,0}(y,z),y,z) + \hat{\psi}_{1,x}(x_{c/r,0}(y,z),y,z)\right] \\ &\quad = \frac{c}{r}, \end{aligned} \tag{2.4.6}$$

$$\hat{\psi}_0(x_{0,0}(y,z),y,z)+\sqrt{\delta}\left[x_{0,1}(y,z)\,\hat{\psi}_{0,x}(x_{0,0}(y,z),y,z)+\hat{\psi}_1(x_{0,0}(y,z),y,z)\right]=1, \tag{2.4.7}$$

and

$$\hat{\psi}_{0,x}(x_{0,0}(y,z),y,z)+\sqrt{\delta}\left[x_{0,1}(y,z)\,\hat{\psi}_{0,xx}(x_{0,0}(y,z),y,z)+\hat{\psi}_{1,x}(x_{0,0}(y,z),y,z)\right]=0. \tag{2.4.8}$$

We begin by approximating $\hat{\psi}_0$ and the free boundaries $x_{c/r,0}$ and $x_{0,0}$. Then, we use the boundaries $x_{c/r,0}$ and $x_{0,0}$ as *fixed* boundaries to determine $\hat{\psi}_1$. As one can see from equations (2.4.5)-(2.4.8), this fixing of the boundaries introduces an $\mathcal{O}(\sqrt{\delta})$-error into $\hat{\psi}_1$ in $\mathcal{O}(\sqrt{\delta})$-neighborhoods of $x_{c/r,0}$ and $x_{0,0}$.

Terms of order δ^0

By inserting (2.4.1)-(2.4.4) into (2.3.4) and collecting terms of order δ^0, we obtain the following free-boundary problem:

$$\begin{cases} cx+\left(\frac{1}{\epsilon}\,\mathcal{L}_0+\frac{1}{\sqrt{\epsilon}}\,\mathcal{L}_1+\mathcal{L}_2\right)\hat{\psi}_0-\frac{1}{\epsilon}\,\nu^2\,(1-\rho_{12}^2)\,\frac{\hat{\psi}_{0,xy}^2}{\hat{\psi}_{0,xx}}=0; \\ \hat{\psi}_0(x_{c/r,0}(y,z),y,z)=\frac{c}{r}\,x_{c/r,0}(y,z), \quad \hat{\psi}_{0,x}(x_{c/r,0}(y,z),y,z)=\frac{c}{r}; \\ \hat{\psi}_0(x_{0,0}(y,z),y,z)=1, \quad \hat{\psi}_{0,x}(x_{0,0}(y,z),y,z)=0. \end{cases} \tag{2.4.9}$$

Terms of order $\sqrt{\delta}$

Similarly, by comparing terms of order $\sqrt{\delta}$ and using $x_{c/r,0}$ and $x_{0,0}$ as fixed boundaries for $\hat{\psi}_1$, we obtain the following boundary-value problem:

$$\begin{cases} \left(\frac{1}{\epsilon}\,\mathcal{L}_0+\frac{1}{\sqrt{\epsilon}}\,\mathcal{L}_1+\mathcal{L}_2\right)\hat{\psi}_1+\left(\mathcal{M}_1+\frac{1}{\sqrt{\epsilon}}\,\mathcal{M}_3\right)\hat{\psi}_0 \\ \quad +\frac{1}{\epsilon}\,\nu^2\,(1-\rho_{12}^2)\left(\left(\frac{\hat{\psi}_{0,xy}}{\hat{\psi}_{0,xx}}\right)^2\hat{\psi}_{1,xx}-2\,\frac{\hat{\psi}_{0,xy}}{\hat{\psi}_{0,xx}}\,\hat{\psi}_{1,xy}\right) \\ \quad -\sqrt{\frac{2}{\epsilon}}\,\nu\,h(z)(\rho_{23}-\rho_{12}\rho_{13})\frac{\hat{\psi}_{0,xy}\,\hat{\psi}_{0,xz}}{\hat{\psi}_{0,xx}}=0; \\ \hat{\psi}_1(x_{c/r,0}(y,z),y,z)=0, \quad \hat{\psi}_1(x_{0,0}(y,z),y,z)=0. \end{cases} \tag{2.4.10}$$

Next, we expand the solutions of (2.4.9) and (2.4.10) in powers of $\sqrt{\epsilon}$:

$$\hat{\psi}_0(x,y,z) = \hat{\psi}_{0,0}(x,y,z) + \sqrt{\epsilon}\,\hat{\psi}_{0,1}(x,y,z) + \epsilon\,\hat{\psi}_{0,2}(x,y,z) + \cdots, \tag{2.4.11}$$

and

$$\hat{\psi}_1(x,y,z) = \hat{\psi}_{1,0}(x,y,z) + \sqrt{\epsilon}\,\hat{\psi}_{1,1}(x,y,z) + \epsilon\,\hat{\psi}_{1,2}(x,y,z) + \cdots. \tag{2.4.12}$$

Similarly, expand the free boundaries $x_{c/r,0}$ and $x_{0,0}$ in powers of $\sqrt{\epsilon}$:

$$x_{c/r,0}(y,z) = x_{c/r,0,0}(y,z) + \sqrt{\epsilon}\,x_{c/r,0,1}(y,z) + \epsilon\,x_{c/r,0,2}(y,z) + \cdots, \tag{2.4.13}$$

and

$$x_{0,0}(y,z) = x_{0,0,0}(y,z) + \sqrt{\epsilon}\,x_{0,0,1}(y,z) + \epsilon\,x_{0,0,2}(y,z) + \cdots. \tag{2.4.14}$$

Substitute (2.4.11) and (2.4.12) into (2.4.9) and (2.4.10), respectively, and collect terms of the same order of $\sqrt{\epsilon}$. As discussed earlier, we determine the free boundaries $x_{c/r,0,0}(y,z)$ and $x_{0,0,0}(y,z)$ via a free-boundary problem for $\hat{\psi}_{0,0}$; then, we use these boundaries as the *fixed* boundaries for $\hat{\psi}_{0,1}$ and $\hat{\psi}_{1,0}$.

Terms of order $1/\epsilon$ in (2.4.9)

By matching terms of order $1/\epsilon$ in (2.4.9), we obtain the following:

$$\mathcal{L}_0\,\hat{\psi}_{0,0} - \nu^2\left(1-\rho_{12}^2\right)\frac{\hat{\psi}_{0,0,xy}^2}{\hat{\psi}_{0,0,xx}} = 0, \tag{2.4.15}$$

or equivalently

$$(m-y)\,\hat{\psi}_{0,0,y} + \nu^2\,\hat{\psi}_{0,0,yy} - \nu^2\left(1-\rho_{12}^2\right)\frac{\hat{\psi}_{0,0,xy}^2}{\hat{\psi}_{0,0,xx}} = 0. \tag{2.4.16}$$

We, therefore, look for an $\hat{\psi}_{0,0}$ independent of y; otherwise, $\hat{\psi}_{0,0}$ will experience exponential growth as y goes to $\pm\infty$ Fouque et al. [43, 44]. We also seek free boundaries $x_{c/r,0,0}$ and $x_{0,0,0}$ independent of y.

Terms of order $1/\sqrt{\epsilon}$ in (2.4.9)

By matching terms of order $1/\sqrt{\epsilon}$ in (2.4.9) and using the fact that $\hat{\psi}_{0,0,y} \equiv 0$, we obtain the following:

$$\mathcal{L}_0 \, \hat{\psi}_{0,1} = 0. \tag{2.4.17}$$

Therefore, we look for an $\hat{\psi}_{0,1}$ independent of y; otherwise, $\hat{\psi}_{0,1}$ will experience exponential growth as y goes to $\pm\infty$.

Terms of order ϵ^0 in (2.4.9)

By matching terms of order ϵ^0 in (2.4.9) and using the fact that $\hat{\psi}_{0,0,y} = \hat{\psi}_{0,1,y} \equiv 0$, we obtain the following Poisson equation (in y) for $\hat{\psi}_{0,2}$:

$$\mathcal{L}_0 \, \hat{\psi}_{0,2} = -cx - \mathcal{L}_2 \, \hat{\psi}_{0,0}. \tag{2.4.18}$$

The solvability condition for this equation requires that $cx + \mathcal{L}_2 \, \hat{\psi}_{0,0}$ be centered with respect to the invariant distribution $\mathcal{N}(m, \nu^2)$ of the Ornstein-Uhlenbeck process Y. Specifically,

$$\left\langle cx + \mathcal{L}_2 \, \hat{\psi}_{0,0} \right\rangle = cx + \langle \mathcal{L}_2 \rangle \, \hat{\psi}_{0,0} = 0, \tag{2.4.19}$$

in which $\langle \cdot \rangle$ denotes averaging with respect to the distribution $\mathcal{N}(m, \nu^2)$:

$$\langle v \rangle = \frac{1}{\sqrt{2\pi\nu^2}} \int_{-\infty}^{\infty} v(y) \, e^{-\frac{(y-m)^2}{2\nu^2}} \, \mathrm{d}y. \tag{2.4.20}$$

In (4.43), the averaged operator $\langle \mathcal{L}_2 \rangle$ is defined by

$$\langle \mathcal{L}_2 \rangle \, v = -\lambda \, v - (r - \lambda) \, x \, v_x + \frac{1}{2} \left(\frac{\mu - r}{\sigma_*(z)} \right)^2 x^2 \, v_{xx}, \tag{2.4.21}$$

in which $\sigma_*(z)$ is given by

$$\frac{1}{\sigma_*^2(z)} = \left\langle \frac{1}{f^2(y, z)} \right\rangle. \tag{2.4.22}$$

Thus, we have the following free-boundary problem for $\hat{\psi}_{0,0}$:

$$\begin{cases} cx - \lambda\,\hat{\psi}_{0,0} - (r-\lambda)\,x\,\hat{\psi}_{0,0,x} + s(z)\,x^2\,\hat{\psi}_{0,0,xx} = 0\,; \\ \hat{\psi}_{0,0}(x_{c/r,0,0}(z),z) = \frac{c}{r}\,x_{c/r,0,0}(z), \quad \hat{\psi}_{0,0,x}(x_{c/r,0,0}(z),z) = \frac{c}{r}\,; \\ \hat{\psi}_{0,0}(x_{0,0,0}(z),z) = 1, \quad \hat{\psi}_{0,0,x}(x_{0,0,0}(z),z) = 0. \end{cases} \tag{2.4.23}$$

with $s(z) = \frac{1}{2}\left(\frac{\mu-r}{\sigma_*(z)}\right)^2$. The general solution of the differential equation in (2.4.23) is given by

$$\hat{\psi}_{0,0}(x,z) = D_1(z)\,x^{B_1(z)} + D_2(z)\,x^{B_2(z)} + \frac{c}{r}\,x, \tag{2.4.24}$$

in which

$$B_1(z) = \frac{1}{2s(z)}\left[(r-\lambda+s(z)) + \sqrt{(r-\lambda+s(z))^2 + 4\lambda s(z)}\right] > 1, \tag{2.4.25}$$

and

$$B_2(z) = \frac{1}{2s(z)}\left[(r-\lambda+s(z)) - \sqrt{(r-\lambda+s(z))^2 + 4\lambda s(z)}\right] < 0. \tag{2.4.26}$$

We determine D_1 and D_2 from the free-boundary conditions.

The free-boundary conditions imply that

$$D_1(z)\,x_{c/r,0,0}(z)^{B_1(z)} + D_2(z)\,x_{c/r,0,0}(z)^{B_2(z)} + \frac{c}{r}\,x_{c/r,0,0}(z) = \frac{c}{r}\,x_{c/r,0,0}(z), \tag{2.4.27}$$

$$D_1(z)\,B_1(z)\,x_{c/r,0,0}(z)^{B_1(z)-1} + D_2(z)\,B_2(z)\,x_{c/r,0,0}(z)^{B_2(z)-1} + \frac{c}{r} = \frac{c}{r}, \tag{2.4.28}$$

$$D_1(z)\,x_{0,0,0}(z)^{B_1(z)} + D_2(z)\,x_{0,0,0}(z)^{B_2(z)} + \frac{c}{r}\,x_{0,0,0}(z) = 1, \tag{2.4.29}$$

and

$$D_1(z)\,B_1(z)\,x_{0,0,0}(z)^{B_1(z)-1} + D_2(z)\,B_2(z)\,x_{0,0,0}(z)^{B_2(z)-1} + \frac{c}{r} = 0, \tag{2.4.30}$$

which gives us four equations to determine the four unknowns D_1, D_2, $x_{c/r,0,0}$, and $x_{0,0,0}$. Indeed, the solution to these equations is

$$D_1(z) = -\frac{1}{B_1(z)-1}\left(\frac{c}{r}\cdot\frac{B_1(z)-1}{B_1(z)}\right)^{B_1(z)}, \tag{2.4.31}$$

$$D_2(z) \equiv 0, \tag{2.4.32}$$

$$x_{c/r,0,0}(z) \equiv 0, \tag{2.4.33}$$

and

$$x_{0,0,0}(z) = \frac{B_1(z)}{B_1(z)-1} \cdot \frac{r}{c}. \tag{2.4.34}$$

It follows that

$$\hat{\psi}_{0,0}(x,z) = -\frac{1}{B_1(z)-1}\left(\frac{c}{r}\cdot\frac{B_1(z)-1}{B_1(z)}\cdot x\right)^{B_1(z)} + \frac{c}{r}\,x. \tag{2.4.35}$$

Terms of order $\sqrt{\epsilon}$ in (2.4.9)

By matching terms of order $\sqrt{\epsilon}$ in (2.4.9) and using the fact that $\hat{\psi}_{0,0,y} = \hat{\psi}_{0,1,y} = 0$, we obtain the following Poisson equation (in y) for $\hat{\psi}_{0,3}$:

$$\mathcal{L}_0\,\hat{\psi}_{0,3} = -\mathcal{L}_1\,\hat{\psi}_{0,2} - \mathcal{L}_2\,\hat{\psi}_{0,1}. \tag{2.4.36}$$

As above, the solvability condition for this equation requires that

$$\left\langle \mathcal{L}_1\,\hat{\psi}_{0,2} + \mathcal{L}_2\,\hat{\psi}_{0,1} \right\rangle = 0, \tag{2.4.37}$$

in which

$$\hat{\psi}_{0,2}(x,z) = \mathcal{L}_0^{-1}\left(-cx - \mathcal{L}_2\,\hat{\psi}_{0,0}\right). \tag{2.4.38}$$

It follows that $\hat{\psi}_{0,1}$ solves

$$\langle \mathcal{L}_2 \rangle\,\hat{\psi}_{0,1} = \left\langle \mathcal{L}_1\mathcal{L}_0^{-1}\left(cx + \mathcal{L}_2\,\hat{\psi}_{0,0}\right)\right\rangle. \tag{2.4.39}$$

Recall that we impose the (fixed) boundary conditions $\hat{\psi}_{0,1}(x_{c/r,0,0}(z),z) = 0$ and $\hat{\psi}_{0,1}(x_{0,0,0}(z),z) = 0$ at $x_{c/r,0,0}(z) \equiv 0$ and $x_{0,0,0}(z) = \frac{B_1(z)}{B_1(z)-1}\cdot\frac{r}{c}$.

From (2.4.38), it is straightforward to show that $\hat{\psi}_{0,2}$ can be expressed as follows:

$$\hat{\psi}_{0,2}(x,y,z) = -D_1(z)\,B_1(z)\,(B_1(z)-1)\,x^{B_1(z)}\,\eta(y,z), \tag{2.4.40}$$

in which η solves

$$(m-y)\eta_y + \nu^2\,\eta_{yy} = \frac{1}{2}\left(\frac{\mu-r}{f(y,z)}\right)^2 - \frac{1}{2}\left(\frac{\mu-r}{\sigma^*(z)}\right)^2 = \frac{1}{2}\left(\frac{\mu-r}{f(y,z)}\right)^2 - s(z). \quad (2.4.41)$$

It follows that the right-hand side of (2.4.39) equals

$$\begin{aligned} &-\rho_{12}\,(\mu-r)\,\nu\sqrt{2}\,D_1(z)\,B_1^2(z)\,(B_1(z)-1)\,x^{B_1(z)}\left\langle \frac{\eta_y(y,z)}{f(y,z)}\right\rangle \\ &= \rho_{12}\,(\mu-r)\,\sqrt{2}\nu\,D_1(z)\,B_1^2(z)\,(B_1(z)-1)\,x^{B_1(z)}\left\langle \tilde{F}(y,z)\left(\frac{1}{2}\left(\frac{\mu-r}{f(y,z)}\right)^2 - s(z)\right)\right\rangle, \end{aligned} \quad (2.4.42)$$

in which $\tilde{F}$ is an antiderivative of $1/f$ with respect to y; that is,

$$\tilde{F}_y(y,z) = \frac{1}{f(y,z)}. \quad (2.4.43)$$

From (2.4.39) and (2.4.42), we obtain that $\hat{\psi}_{0,1}$ equals

$$\hat{\psi}_{0,1}(x,z) = \tilde{D}_1(z)\,x^{B_1(z)} + \tilde{D}_2(z)\,x^{B_2(z)} + A(z)\,x^{B_1(z)}\,\ln x, \quad (2.4.44)$$

in which B_1 and B_2 are given in (2.4.25) and (2.4.26), respectively, and A is given by

$$A(z) = \frac{\rho_{12}\,(\mu-r)\,\sqrt{2}\nu\,D_1(z)\,B_1^2(z)\,(B_1(z)-1)}{(2B_1(z)-1)\,s(z) - (r-\lambda)}\left\langle \tilde{F}(y,z)\left(\frac{1}{2}\left(\frac{\mu-r}{f(y,z)}\right)^2 - s(z)\right)\right\rangle. \quad (2.4.45)$$

The functions $\tilde{D}_1$ and $\tilde{D}_2$ are given by the (fixed) boundary conditions at $x_{c/r,0,0}(z) \equiv 0$ and $x_{0,0,0}(z) = \frac{B_1(z)}{B_1(z)-1}\cdot\frac{r}{c}$, from which it follows that

$$\begin{aligned} \hat{\psi}_{0,1}(x,z) &= A(z)\,x^{B_1(z)}\left(\ln x - \ln\left(\frac{B_1(z)}{B_1(z)-1}\cdot\frac{r}{c}\right)\right) \\ &= A(z)\,x^{B_1(z)}\,\ln\left(x\cdot\frac{B_1(z)-1}{B_1(z)}\cdot\frac{c}{r}\right). \end{aligned} \quad (2.4.46)$$

Next, we focus on (2.4.10) to find $\hat{\psi}_{1,0}$, after which we will approximate $\hat{\psi}$ by $\hat{\psi}_{0,0} + \sqrt{\epsilon}\,\hat{\psi}_{0,1} + \sqrt{\delta}\,\hat{\psi}_{1,0}$.

Terms of order $1/\epsilon$ in (2.4.10)

By matching terms of order $1/\epsilon$ in (2.4.10), we obtain the following:

$$\mathcal{L}_0\,\hat{\psi}_{1,0} = 0, \tag{2.4.47}$$

from which it follows that $\hat{\psi}_{1,0}$ is independent of y; otherwise, $\hat{\psi}_{1,0}$ will experience exponential growth as y goes to $\pm\infty$ Fouque et al. [44].

Terms of order $1/\sqrt{\epsilon}$ in (2.4.10)

By matching terms of order $1/\sqrt{\epsilon}$ in (2.4.10) and using the fact that $\hat{\psi}_{1,0,y} \equiv 0$, we obtain the following:

$$\mathcal{L}_0\,\hat{\psi}_{1,1} = 0. \tag{2.4.48}$$

Therefore, we look for an $\hat{\psi}_{1,1}$ independent of y; otherwise, $\hat{\psi}_{1,1}$ will experience exponential growth as y goes to $\pm\infty$.

Terms of order ϵ^0 in (2.4.10)

By matching terms of order ϵ^0 in (2.4.10) and using the fact that $\hat{\psi}_{1,0,y} = \hat{\psi}_{1,1,y} \equiv 0$, we obtain the following Poisson equation (in y) for $\hat{\psi}_{1,2}$:

$$\mathcal{L}_0\,\hat{\psi}_{1,2} = -\mathcal{L}_2\,\hat{\psi}_{1,0} + \rho_{13}\,\frac{\mu - r}{f(y,z)}\,h(z)\,x\,\hat{\psi}_{0,0,xz}. \tag{2.4.49}$$

The solvability condition for this equation requires that

$$\left\langle -\mathcal{L}_2\,\hat{\psi}_{1,0} + \rho_{13}\,\frac{\mu - r}{f(y,z)}\,h(z)\,x\,\hat{\psi}_{0,0,xz} \right\rangle = 0, \tag{2.4.50}$$

or equivalently,

$$\langle \mathcal{L}_2 \rangle\,\hat{\psi}_{1,0} = \rho_{13}\left\langle \frac{\mu - r}{f(y,z)} \right\rangle h(z)\,x\,\hat{\psi}_{0,0,xz}. \tag{2.4.51}$$

with boundary conditions $\hat{\psi}_{1,0}(x_{c/r,0,0}(z),z) = 0$ and $\hat{\psi}_{1,0}(x_{0,0,0}(z),z) = 0$ at the boundaries $x_{c/r,0,0}(z) \equiv 0$ and $x_{0,0,0}(z) = \frac{B_1(z)}{B_1(z)-1} \cdot \frac{r}{c}$. It follows that $\hat{\psi}_{1,0}$ is given by

$$\hat{\psi}_{1,0}(z) = x^{B_1(z)} \ln\left(x \cdot \frac{B_1(z)-1}{B_1(z)} \cdot \frac{c}{r}\right) \left[A_1(z) + A_2(z) \ln\left(x \cdot \frac{B_1(z)}{B_1(z)-1} \cdot \frac{r}{c}\right)\right], \tag{2.4.52}$$

in which A_1 and A_2 are

$$A_1(z) = \frac{H_1(z)}{(2B_1(z)-1)\,s(z) - (r-\lambda)} - \frac{H_2(z)\,s(z)}{\left[(2B_1(z)-1)\,s(z) - (r-\lambda)\right]^2}, \tag{2.4.53}$$

and

$$A_2(z) = \frac{1}{2} \cdot \frac{H_2(z)}{(2B_1(z)-1)\,s(z) - (r-\lambda)}, \tag{2.4.54}$$

with H_1 and H_2 functions of z defined by

$$\begin{aligned} &H_1(z) + H_2(z)\,\ln x \\ &= -\rho_{13}\,h(z) \left\langle \frac{\mu - r}{f(y,z)} \right\rangle \frac{B_1'(z)}{B_1(z)-1} \left(\frac{B_1(z)-1}{B_1(z)} \cdot \frac{c}{r}\right)^{B_1(z)} \left[1 + B_1(z)\,\ln\left(x \cdot \frac{B_1(z)-1}{B_1(z)} \cdot \frac{c}{r}\right)\right]. \end{aligned} \tag{2.4.55}$$

2.4.1 The approximation of the probability of lifetime ruin and the optimal investment strategy

Combining (2.4.35), (2.4.46), and (2.4.52), we obtain the following approximation of $\hat{\psi}$

$$\begin{aligned} \hat{\psi}^{\epsilon,\delta}(x,z) &= \hat{\psi}_{0,0}(x,z) + \sqrt{\epsilon}\,\hat{\psi}_{0,1}(x,z) + \sqrt{\delta}\,\hat{\psi}_{1,0}(x,z) \\ &= -\frac{1}{B_1(z)-1}\left(\frac{c}{r} \cdot \frac{B_1(z)-1}{B_1(z)} \cdot x\right)^{B_1(z)} + \frac{c}{r}\,x \\ &\quad + \sqrt{\epsilon}\,A(z)\,x^{B_1(z)} \ln\left(x \cdot \frac{B_1(z)-1}{B_1(z)} \cdot \frac{c}{r}\right) \\ &\quad + \sqrt{\delta}\,x^{B_1(z)} \ln\left(x \cdot \frac{B_1(z)-1}{B_1(z)} \cdot \frac{c}{r}\right) \left[A_1(z) + A_2(z) \ln\left(x \cdot \frac{B_1(z)}{B_1(z)-1} \cdot \frac{r}{c}\right)\right], \end{aligned} \tag{2.4.56}$$

in which A, A_1, and A_2, are specified in (2.4.45), (2.4.53), and (2.4.54), respectively.

We also approximate the dual of the optimal investment strategy up to the first powers of $\sqrt{\epsilon}$ and $\sqrt{\delta}$, as we did for $\hat{\psi}$. Using (2.3.25), we obtain

$$\hat{\pi}^{\epsilon,\delta}(x,z) = -\frac{\mu - r}{f^2(y,z)}\, x\, \hat{\psi}_{0,0,xx} + \sqrt{\epsilon}\left(-\frac{\mu - r}{f^2(y,z)}\, x\, \hat{\psi}_{0,1,xx} + \rho_{12}\, \frac{\nu\sqrt{2}}{f(y,z)}\, \hat{\psi}_{0,2,xy}\right) + \sqrt{\delta}\left(-\frac{\mu - r}{f^2(y,z)}\, x\, \hat{\psi}_{1,0,xx} + \rho_{13}\, \frac{h(z)}{f(y,z)}\, \hat{\psi}_{0,0,xz}\right). \tag{2.4.57}$$

Given $w \in \mathbb{R}_+$, we solve for x using $w = \hat{\psi}^{\epsilon,\delta}_x(x,z)$. Then, we let $\psi^{\epsilon,\delta}(w,z) := \hat{\psi}^{\epsilon,\delta}(x,z) - xw$, thereby performing the calculation in equation (2.3.13). We also denote by $\pi^{\epsilon,\delta}$ the function that satisfies $\pi^{\epsilon,\delta}(w,z) := \hat{\pi}^{\epsilon,\delta}(x,z)$. Note that the resulting approximation of ψ is not guaranteed to be a probability; however, this is not a problem in the numerical experiments we consider in the next section.

2.5 Numerical Solution using the Markov Chain Approximation Method

In this section, we describe how to construct a numerical algorithm for the original optimal control problem directly using the Markov Chain Approximation Method (MCAM); see e.g. Kushner and Dupuis [70], Kushner [69]. For the ease of presentation, we will describe the numerical algorithm only when the fast scale volatility factor is present. In what follows ρ will denote the correlation between the Brownian motion driving the stock and the one driving the fast factor, that is, $\rho = \rho_{12}$.

Let us fix an h-grid, that is, a rectangular compact domain $G^h \subset \mathbf{R}^2$ with the same spacing h in both directions. We choose an initial guess (on this grid) for a candidate optimal strategy. Denote this strategy by π. Then, our goal is to create a discrete-time Markov chain $(\xi^h_n)_{n\geq 0}$ that lives on G^h and that satisfies the local

consistency condition

$$\begin{aligned}\mathbb{E}^{h,\pi}_{x,n}[\Delta\xi^h_{n+1}] &= b(x,\pi)\Delta t^{\pi,h}(x,\pi) + o(\Delta t^h),\\ \mathrm{Cov}^{h,\pi}_{x,n}[\Delta\xi^h_{n+1}] &= A(x,\pi)\Delta t^{\pi,h}(x,\pi) + o(\Delta t^h),\end{aligned} \tag{2.5.1}$$

in which $\Delta\xi_{n+1} = \xi_{n+1} - \xi_n$, and b and A denote the drift and the covariance of the vector $X_t = (W_t, Y_t)$, respectively. (The Markov chain is constructed to approximate this vector in a certain sense.) $\mathbb{E}^{h,\pi}_{x,n}$ denotes the expectation, given that the state of the Markov chain at time n is x. In (2.5.1) the quantity Δt^h (called the interpolation interval) is to be chosen so that it goes to zero as $h \to 0$. We also do not want this quantity to depend on the state variables or the control variable.

Since G^h is a compact domain, we impose reflecting boundary conditions at its edges. (Natural boundaries exist for $W(t)$, specifically 0 and the safe level $\frac{c}{r}$. However, Y_t lives on an infinite region.) For example, we choose the transition probabilities to be $p^{\pi,h}((w,y),(w,y-h)) = 1$, when y is as large it can be in G^h and for all $w \in [0, \frac{c}{r}]$.

2.5.1 Constructing the approximating Markov Chain

When $\rho = 0$.

Denote $\alpha = \frac{1}{\epsilon}, \beta = \nu\sqrt{\frac{2}{\epsilon}}$. We obtain the transition probabilities of the Markov chain ξ^h as

$$\begin{cases} p^{\pi,h}((w,y),(w,y\pm h)) &= \dfrac{\beta^2/2 + h\alpha(m-y)^{\pm}}{\widetilde{Q}^h}\,,\\ p^{\pi,h}((w,y),(w\pm h,y)) &= \dfrac{(f(y)\pi(w,y))^2/2 + h(\mu-r)\pi(w,y)^{\pm} + h(rw-c)^{\pm}}{\widetilde{Q}^h}\,,\\ p^{\pi,h}((w,y),(w,y)) &= \frac{\widetilde{Q}^h - Q^{\pi,h}(w,y)}{\widetilde{Q}^h}\,, \end{cases} \tag{2.5.2}$$

and choose the interpolation interval to be

$$\Delta t^h = \frac{h^2}{\widetilde{Q}^h},$$

in which

$$Q^{\pi,h}(w,y) = (\pi f(y))^2 + \beta^2 + h|\alpha(m-y)| + h|(\mu - r)\pi(w,y)| + h|rw - c|,$$

and

$$\widetilde{Q}^h = \max_{(w,y,\pi)} Q^{\pi,h}(w,y),$$

in order to satisfy the local consistency condition. Here $a^{\pm} = \max\{0, \pm a\}$.

When $\rho \neq 0$.

In this case a convenient transition probability matrix solving the local consistency condition is

$$\begin{cases} p^{\pi,h}((w,y),(w,y \pm h)) = \dfrac{(1-\rho^2)\beta^2/2 - |\rho\pi(w,y)|\beta f(y)/2 + h\alpha(m-y)^{\pm}}{\widetilde{Q}^h}, \\ p^{\pi,h}((w,y),(w \pm h,y)) = \dfrac{(f(y)\pi(w,y))^2 - |\rho\pi(w,y)|\beta f(y)}{2\widetilde{Q}^h} \\ \qquad\qquad + \dfrac{h(\mu-r)\pi(w,y)^{\pm} + h(rw-c)^{\pm}}{\widetilde{Q}^h}, \\ p^{\pi,h}((w,y),(w+h,y+h)) = p^{\pi,h}((w,y),(w-h,y-h)) = \dfrac{(\rho\pi(w,y))^{+}\beta f(y)}{2\widetilde{Q}^h}, \\ p^{\pi,h}((w,y),(w+h,y-h)) = p^{\pi,h}((w,y),(w-h,y+h)) = \dfrac{(\rho\pi(w,y))^{-}\beta f(y)}{2\widetilde{Q}^h} \\ p^{\pi,h}((w,y),(w,y)) = \dfrac{\widetilde{Q}^h - Q^{\pi,h}(w,y)}{\widetilde{Q}^h}, \end{cases} \tag{2.5.3}$$

where

$$Q^{\pi,h}(w,y) = (\pi f(y))^2 + \beta^2 - |\rho\pi(w,y)|\beta f(y) + h|\alpha(m-y)| + h|(\mu-r)\pi(w,y)| + h|rw-c|.$$

For values of $|\rho|$ close to 1, the transition probabilities may be negative. The positiveness of these probabilities is equivalent to the diagonal dominance of the

covariance matrix $A = (a_{ij})$. (Recall that we call A diagonally dominant if $a_{ii} - \sum_{j,j\neq i} |a_{ij}| > 0, \quad \forall i.$) The construction of an approximating Markov chain when some of the expressions in (2.5.3) are negative will be discussed next.

When $\rho = 1$ and some of the transition probabilities in (2.5.3) are negative.

We accomplish the construction of the approximating Markov chain in two steps, following Kushner [69]:

(i) Decomposition. As in Kushner and Dupuis [70] Sections 5.3 and 5.4, we decompose X into separate components and build approximating Markov chains to match each component. Then, we combine the transition probabilities appropriately to obtain the approximating Markov chain for X itself.

Let $X = X^{(1)} + X^{(2)}$, in which

$$\mathrm{d}X_t^{(1)} = \begin{pmatrix} \pi f(y) \\ \beta \end{pmatrix} \mathrm{d}B_t^1, \tag{2.5.4}$$

$$\mathrm{d}X_t^{(2)} = \begin{pmatrix} rW_t - c + (\mu - r)\pi_t \\ \alpha(m - Y_t) \end{pmatrix} dt. \tag{2.5.5}$$

Since $\rho = 1$, we take $B^1 = B^2$. Suppose that the form of the locally consistent (with dynamics of $X^{(1)}$ and $X^{(2)}$, respectively) transition probabilities and interpolation intervals are

$$p_1^{\pi,h}(x,\bar{x}) = \frac{n_1^{\pi,h}(x,\bar{x})}{\widetilde{Q}_1^h}\ ,\ \Delta t_1^{\pi,h} = \frac{h^2}{\widetilde{Q}_1^h}\ ,$$
$$p_2^{\pi,h}(x,\bar{x}) = \frac{n_2^{\pi,h}(x,\bar{x})}{\widetilde{Q}_2^h}\ ,\ \Delta t_2^{\pi,h} = \frac{h}{\widetilde{Q}_2^h}\ ,$$

for some $n_1^{\pi,h}(x,\bar{x}), n_2^{\pi,h}(x,\bar{x})$, and appropriate normalizers $\widetilde{Q}_1^h$, $\widetilde{Q}_2^h$. Then, the following transition probabilities and the interpolation interval are locally consistent

with the dynamics of X

$$p^{\pi,h}(x,\bar{x}) = \frac{n_1^{\pi,h}(x,\bar{x}) + h n_2^{\pi,h}(x,\bar{x})}{\widetilde{Q}_1^h + h\widetilde{Q}_2^h}, \quad \Delta t^{\pi,h} = \frac{h^2}{\widetilde{Q}_1^h + h\widetilde{Q}_2^h}. \tag{2.5.6}$$

Since it is easier, we first provide the expression for $p_2^{\pi,h}$:

$$\begin{cases} p_2^{\pi,h}((w,y),(w,y\pm h)|\pi) & = \dfrac{\alpha(m-y)^{\pm}}{\widetilde{Q}_2^h}, \\ p_2^{\pi,h}((w,y),(w\pm h,y)|\pi) & = \dfrac{(\mu-r)\pi(w,y)^{\pm} + (rw-c)^{\pm}}{\widetilde{Q}_2^h}, \\ p_2^{\pi,h}((w,y),(w,y)) & = \dfrac{\widetilde{Q_2}^h - Q_2^{\pi,h}(w,y)}{\widetilde{Q}_2^h}, \end{cases} \tag{2.5.7}$$

where

$$Q_2^{\pi,h}(w,y) = \alpha|m-y| + (\mu-r)|\pi(w,y)| + |rw-c|.$$

The computation of $p_1^{\pi,h}$ is more involved. This is the subject of the next step.

(ii) Variance control. System (2.5.4) is fully degenerate; that is, the corresponding covariance matrix A is not diagonally dominant. The previous technique for building a Markov chain does not work. Instead, we will build an approximating Markov chain by allowing the local consistency condition to be violated by a small margin of error.

If $(\sigma_1,\sigma_2) = (qk_1, qk_2)$ for some constant q and integers k_1, k_2, we could let the transition probability to be $p^h(x, x\pm(hk_1,hk_2)) = 1/2$ and the interpolation interval to be $\Delta t^h = h^2/q^2$, and we would obtain a locally consistent Markov chain. This is not possible in general. For an arbitrary vector (σ_1,σ_2), we can find a pair of integers $k_1(x,\pi), k_2(x,\pi)$, and a real number $\gamma(x,\pi)\in[0,1]$, such that

$$\begin{pmatrix} \sigma_1(x,\pi) \\ \\ \sigma_2 \end{pmatrix} = q(x,\pi) \begin{pmatrix} k_1(x,\pi) \\ \\ k_2(x,\pi) + \gamma(x,\pi) \end{pmatrix}.$$

Since the Markov chain is constrained to the grid G^h, we can only approximately

let it move in the direction of $(\sigma_1, \sigma_2)^T$. We choose

$$\begin{aligned} p^{\pi,h}(x, x \pm h(k_1, k_2)^T) &= p_1/2, \\ p^{\pi,h}(x, x \pm h(k_1, k_2+1)^T) &= p_2/2, \end{aligned} \tag{2.5.8}$$

in which $p_1 + p_2 = 1$, and p_1 and p_2 will be appropriately chosen in what follows. The mean and the covariance of the approximating chain is

$$\begin{aligned} \mathbb{E}^{h,\pi}_{x,\pi}[\Delta\xi^h(x,\pi)] &= 0, \\ \mathbb{E}^{h,\pi}_{x,\pi}[\Delta\xi^h(x,\pi)\Delta\xi^h(x,\pi)^T] &= h^2 C(x,\pi), \end{aligned} \tag{2.5.9}$$

where

$$\begin{aligned} C(x,\pi) &= p_1 \begin{pmatrix} k_1^2 & k_1 k_2 \\ k_1 k_2 & k_2^2 \end{pmatrix} + p_2 \begin{pmatrix} k_1^2 & k_1(k_2+1) \\ k_1(k_2+1) & (k_2+1)^2 \end{pmatrix} \\ &= \begin{pmatrix} k_1^2 & k_1(k_2+p_2) \\ k_1(k_2+p_2) & k_2^2 + 2pk_2 + p_2 \end{pmatrix}. \end{aligned} \tag{2.5.10}$$

We choose the interpolation interval to be $\Delta t^{\pi,h}(x,\pi) = h^2/q^2$. On the other hand

$$a(x,\pi) = A(x,\pi)/q^2 = \begin{pmatrix} k_1^2 & k_1(k_2+\gamma) \\ k_1(k_2+\gamma) & (k_2+\gamma)^2 \end{pmatrix},$$

and we see that if we pick $p_2 = \gamma$, then $C_{11} = a_{11}$ and $C_{12} = a_{12}$ match, but we violate the local consistency condition by

$$\frac{C_{22} - a_{22}^2}{a_{22}^2} = \frac{\gamma(1-\gamma)}{(k_2+\gamma)^2} = O\left(\frac{1}{k_2^2}\right). \tag{2.5.11}$$

We will choose k_2 sufficiently large so that the local consistency condition is almost satisfied, and the numerical noise in (2.5.11) is significantly reduced.

The case when $\rho \in (-1,1)$ and some of the transition probabilities in (2.5.3) are negative

We will decompose the state variable into three components:

$$\mathrm{d}\vec{X}_t = \begin{pmatrix} \mathrm{d}W_t \\ \mathrm{d}Y_t \end{pmatrix} = \begin{pmatrix} rW_t - c + (\mu - r)\pi_t \\ \alpha(m - Y_t) \end{pmatrix} dt + \begin{pmatrix} \pi_t f(Y_t) \\ \beta\rho \end{pmatrix} \mathrm{d}B_t^1 + \begin{pmatrix} 0 & 0 \\ 0 & \beta\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \mathrm{d}B_t^1 \\ \mathrm{d}B_t^2 \end{pmatrix}, \tag{2.5.12}$$

that is, a drift component, a fully degenerate noise component, and a noise component with diagonally dominated covariance matrix. We can build an approximating Markov chain for each component separately and then combine them as discussed above.

2.5.2 Approximating the probability of ruin and updating the strategy

We solve the system of linear equations

$$V^{\pi,h}(x) = e^{-\lambda \Delta t^{\pi,h}} \sum_{\tilde{x} \in G^h} p^{\pi,h}(x, \tilde{x}) V^{\pi,h}(\tilde{x}), \tag{2.5.13}$$

with boundary conditions $V^{\pi,h}(0, y) = 1$ and $V^{\pi,h}(c/r, y) = 0$. This is the dynamic programming equation for a probability of ruin problem when the underlying state variable is the Markov chain ξ^h. In the next step, we update our candidate for the optimal strategy. For convenience, denote $V^{\pi,h}$ by V.

In the interior points of the grid

$$\begin{aligned}\pi(w,y) = &- \frac{h(\mu - r)[V(w+h,y) - V(w,y)]}{f^2(y)[V(w+h,y) + V(w-h,y) - 2V(w,y)]} \\ &+ \frac{(\beta/2)\rho f(y)\left[V(w+h,y+h) + V(w,y-h) - V(w+h,y-h) - V(w,y+h)\right]}{f^2(y)\left[V(w+h,y) + V(w-h,y) - 2V(w,y)\right]}.\end{aligned}$$

On the wealth dimension boundaries of the grid, we let $\pi(c/r, y) = 0$ and

$$\begin{aligned}\pi(0,y) = &- \frac{h(\mu - r)[V(h,y) - V(0,y)]}{f^2(y)\left[2V(0,y) - 5V(h,y) + 4V(2h,y) - V(3h,y)\right]} \\ &+ \frac{(\beta/2)\rho f(y)\left[V(h,y+h) + V(0,y-h) - V(h,y-h) - V(0,y+h)\right]}{f^2(y)\left[2V(0,y) - 5V(h,y) + 4V(2h,y) - V(3h,y)\right]}.\end{aligned}$$

The updates of the optimal strategy for the maximum and minimum values of y are similar.

Iteration

Once the optimal strategy is updated, we go back and update the transition probabilities and solve the system of linear equations in (2.5.13) to update the value

function. This iteration continues until the improvement in the value function is smaller than an exogenously picked threshold.

Two Technical Issues

- The initial guess of the optimal strategy is important. For $\rho = 0$, we take the initial strategy as the one in constant volatility case, where the closed-form solution is available in Young [97]. For $\rho \neq 0$, we take the final strategy computed from zero-correlation case ($\rho = 0$) as the initial guess. This initial guess makes the algorithm converge fast.

- For $\rho \neq 0$, the covariance matrix of the wealth process and volatility factor, in general, does not satisfy the diagonal dominance condition. The problem is more serious for the slow factor, since its variance is of the order of δ, and the numerical noise using "'variance control"' is far greater. To solve this issue we perform a "scale adjustment" to increase the variance of the factor. For example, if we define $\bar{Z}_t = 100 Z_t$, then the dynamic of the system becomes

$$\begin{aligned} \frac{\mathrm{d}S_t}{S_t} &= \mu dt + f\left(Y_t, \frac{\bar{Z}_t}{100}\right) \mathrm{d}B_t^1, \\ \mathrm{d}\bar{Z}_t &= \delta(100m - \bar{Z}_t)dt + 100\sqrt{\delta}\sqrt{2}\nu_2 \mathrm{d}B_t^{(3)}. \end{aligned} \tag{2.5.14}$$

  when $g = (m - z)$ and $h = \sqrt{2}\nu$. The new system is mathematically equivalent to the original one, but with a much bigger variance; thus, the numerical noise in variance control is much smaller. Note that the scheme here is equivalent to choosing a different grid sizes for the volatility and wealth dimensions.

2.6 Numerical Experiments

In order to conduct our numerical experiments we will take the dynamics of the slow factor in (2.2.7) to be

$$\mathrm{d}Z_t = \delta(m - Z_t)dt + \sqrt{\delta}\sqrt{2}\nu \mathrm{d}B_t^{(3)}, \qquad Z_0 = z.$$

We let $f(y,z) = \exp(-y)$ or $f(y,z) = \exp(-z)$ in (2.2.2), depending on whether we want to account for the fast volatility factor or the slow volatility factor in our modeling. We will call $1/\epsilon$ or δ the speed of mean reversion. We will take the correlations between the Brownian motions driving the volatility factors and the stock price to be $\rho = \rho_{13} = \rho_{12}$.

The following parameters are fixed throughout this section:

- $r = 0.02$; the risk-free interest rate is 2% over inflation.
- $\mu = 0.1$; the expected return of risky asset is 10% over inflation.
- $c = 0.1$; the individual consumes at a constant rate of 0.1 unit of wealth per year.
- $\lambda = 0.04$; the hazard rate (force of mortality) is constant such that the expected future lifetime is always 25 years.
- $m = 1.364$ and $\nu = 0.15$, so that the harmonic average volatility, which we will denote by $\sigma_m = \sqrt{1/\mathbb{E}[1/f^2(Y)]} = \sqrt{1/\mathbb{E}[e^{2Y}]} = e^{-m-\nu^2} = 0.25$, in which Y is a normal random variable with mean m and variance ν^2. The distribution of this random variable is the stationary distribution of the process $(Y_t)_{t\geq 0}$; see (2.4.22). [Note that σ_m is very close in value to $\mathbb{E}[f(Y)] = \mathbb{E}[e^{-Y}] = e^{-m+\nu^2/2} = 0.26$.]

In our numerical procedure we use a bounded region for Y and impose reflecting boundary conditions. However, $f(Y_t)$ is not bounded and not bounded away from

zero. On the other hand, the invariant distribution of the process Y is normal with mean 1.364, and variance 0.15^2. So it is with very small probability that Y_t is negative or very large. Therefore, the fact that $f(Y_t)$ is not bounded or bounded away from zero does not affect the accuracy of our numerical work in a significant way.

Observation 1

We give a three-dimensional graph of the minimum probability of ruin and the optimal investment strategy in Figure 2.1, which are computed using MCAM. Here the speed of mean reversion is 0.5, $\rho = 0$, and only one factor is used. In our experiments we observed that the optimal strategy π^* is positive (no-shortselling). As expected we observe that $w \to \psi(w, y)$ is convex and decreasing. Note that $f(y) \to \psi(w, y)$ is increasing. Also, $f(y) \to \pi^*(w, y)$ is decreasing; however, it is not necessarily true that $w \to \pi^*(w, y)$ is decreasing. The latter behavior depends on the value of y.

The probability of ruin does not depend on the sign of the correlation, ρ, between the Brownian motions driving the stock and the one driving the volatility. The larger the magnitude of ρ, the larger the probability of ruin. However, the minimum probability of ruin is quite insensitive to the changes in ρ; see Figure 2.2.

Observation 2

We compare the optimal investment strategy $\pi^*(w, y, z)$ in (2.3.14) to

$$\widetilde{\pi}(w; \sigma) = \frac{\mu - r}{\sigma^2} \frac{c - rw}{(p-1)r} \quad ,$$

in which

$$p = \frac{1}{2r} \left[(r + \lambda + s) + \sqrt{(r + \lambda + s)^2 - 4r\lambda} \right] ,$$

and

$$s = \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2.$$

When we want to emphasize the dependence on σ, we will refer to p as $p(\sigma)$. Young [97] showed that the strategy $\widetilde{\pi}$ is optimal when the volatility is fixed to be σ.

If only the fast factor is present and the speed of mean reversion is 250 ($\epsilon = 0.004$), then $\hat{\psi}^{\epsilon,\delta}$ in (2.4.56) can be expressed as

$$\hat{\psi}^{\epsilon}(w) = \hat{\phi}_{0,0}(x)$$

whose inverse Legendre transform is

$$\psi^{\epsilon}(w;\sigma_m) = \left(1 - \frac{r}{c}w\right)^{p(\sigma_m)}, \tag{2.6.1}$$

which is exactly the minimal probability of ruin if the volatility were fixed at σ_m. Therefore, it is not surprising that for very small values of ϵ, the minimum probability of ruin $\psi(w,y)$, calculated using MCAM, can be approximated by (2.6.1); see Figure 2.3-a. In our numerical calculations and in (2.6.1), we observe that the minimum probability of ruin ψ does not depend on its second variable. This result is intuitive, since when only the fast factor is present whatever the initial value of σ_0 is, the volatility quickly approaches its equilibrium distribution (which is normal with mean σ_m). In fact $\pi_0(w;\sigma_m)$ practically coincides with the optimal investment strategy π^*, which is computed using MCAM; see Figure 2.3-b.

The most important conclusion from Figure 2.3-b is that it is not necessarily true that the optimal investment strategy when there is stochastic volatility is more or less than the optimal investment strategy when the volatility is constant. Comparing $\widetilde{\pi}(w;\sigma)$ and $\pi^*(w,-\ln(\sigma))$ for different values of σ, we see that $\pi^*(w,-\ln(\sigma)) < \widetilde{\pi}(w;\sigma)$ for larger values of σ, whereas the opposite inequality holds for smaller values of σ. The investment amount decreases significantly as the volatility increases.

If only the slow factor is present and the speed of mean reversion is 0.02, then

$$\psi^{\delta}(w,z) = \left(1 - \frac{r}{c}w\right)^{p(e^{-z})}, \tag{2.6.2}$$

approximates the minimum probability of ruin $\psi(w,z)$, which we calculate using MCAM, quite well; compare $\psi(w,-\ln(\sigma))$ and $\psi^{\delta}(w,-\ln(\sigma))$ for different values of σ in Figure 2.4-a. We also compare $\widetilde{\pi}(w;\sigma)$ and $\pi^{*}(w,-\ln(\sigma))$ for several values of σ and draw the same conclusions as before. Also note that the optimal investment strategy is not necessarily a decreasing function of wealth.

When we take the speed of mean reversion to be 0.2 (medium speed), then the probability of ruin starts diverting from what (2.6.1) or (2.6.2) describes; see Figure 2.5-a. As to the comparison of the optimal investment strategy with $\widetilde{\pi}(w;\sigma)$, the same conclusions can be drawn; see Figure 2.5-b.

Observation 3

We compare the performance of several investment strategies in the stochastic volatility environment. Let σ_0 be the initial volatility. We denote by π^{M} the strategy when one only invests in the money market. The corresponding probability of ruin can be explicitly computed as $\psi^{M}(w) = (1 - c/rw)^{1\vee[\lambda/r]}$. We will also denote $\pi^{a}(w) = \widetilde{\pi}(w;\sigma_0)$, $\pi^{b} = \widetilde{\pi}(w;\sigma_m)$, and

$$\pi^{c}(w,y,z) = \frac{\mu - r}{f^2(y,z)}\frac{c - rw}{(p-1)r}. \tag{2.6.3}$$

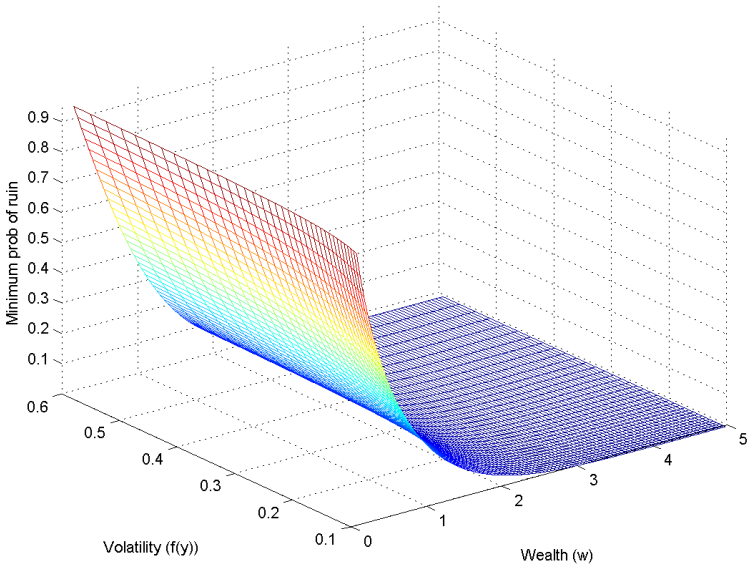
Let $\pi^{\epsilon}(w)$ denote the approximation to the optimal strategy we obtained in Section 2.4.1 when we only use the ϵ-perturbation. Similarly, let $\pi^{\delta}(w,z)$ be the approximation to the optimal strategy when we only use the δ-perturbation.

We obtain the probability of ruin corresponding to a given strategy π by solving the linear partial differential equation $\mathcal{D}^{\pi}v = 0$, (see (2.3.24) for the definition of the

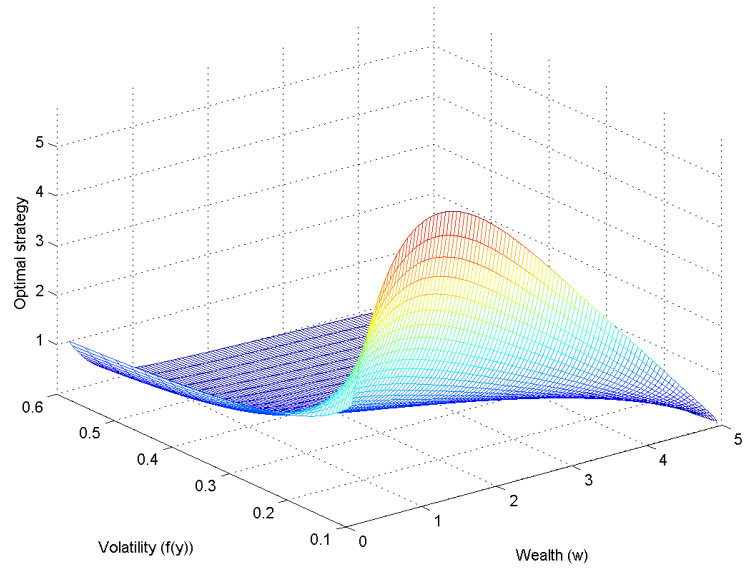
differential operator $\mathcal{D}^{\pi}$) with boundary conditions $v(0, y, z) = 0$ and $v(c/r, y, z) = 1$. (This computation uses the MCAM without iterating.)

In Figure 2.6-2.8 we observe that the performance of π^c and π^{ϵ} are almost as good as the optimal strategy π^*. (Here we are considering a medium mean reversion speed. When the mean reversion speed is much smaller, then π^{δ} would be a better investment strategy.) Moreover, their performances are robust, in that, they do not depend on the initial volatility σ_0. This should be contrasted to π^a and π^b. The former performs relatively well when σ_0 is small, whereas the latter performs better when σ_0 is large. When $\sigma_0 = \sigma_m$, all strategies perform as well as the optimal strategy. Also, observe that for wealthy or very poor individuals the choice of the strategy does not matter as long as they invest in the stock market. The difference is for the individuals who lie in between.

As a result, we conclude that if the individual wants to minimize her probability of ruin in a stochastic volatility environment, she can still use the investment that is optimal for the constant volatility environment. She simply needs to update the volatility in that formula whenever the volatility changes significantly.

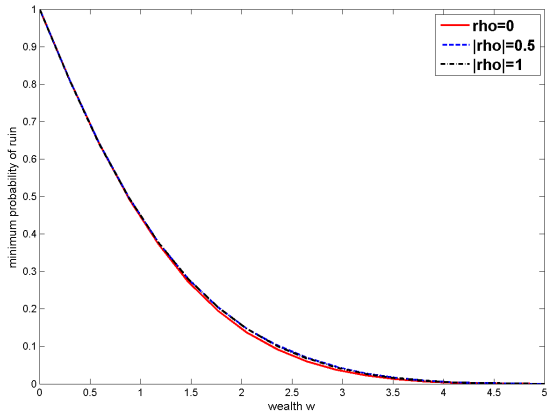


(a) Minimum probability of ruin

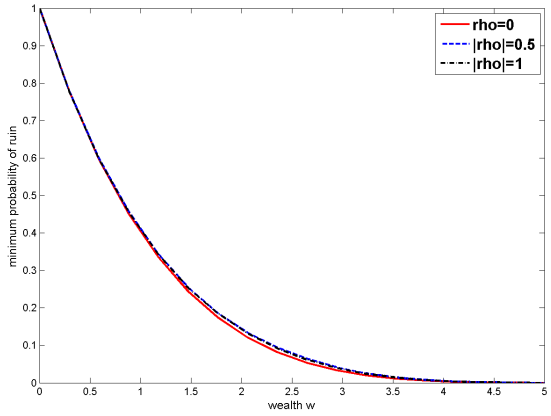


(b) Optimal investment strategy

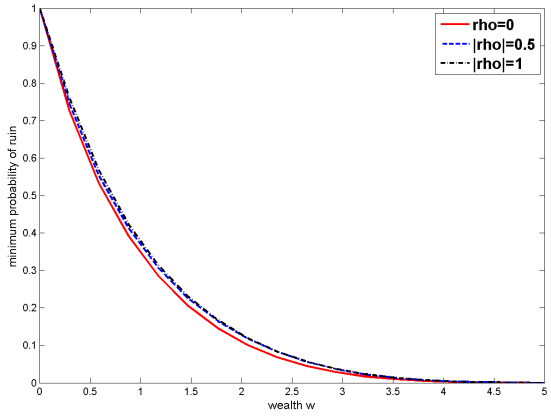
Figure 2.1: Minimum probability of ruin and optimal strategy computed by MCAM. Speed of mean reversion= 0.5.



(a) Probability of ruin at σ_0=0.6

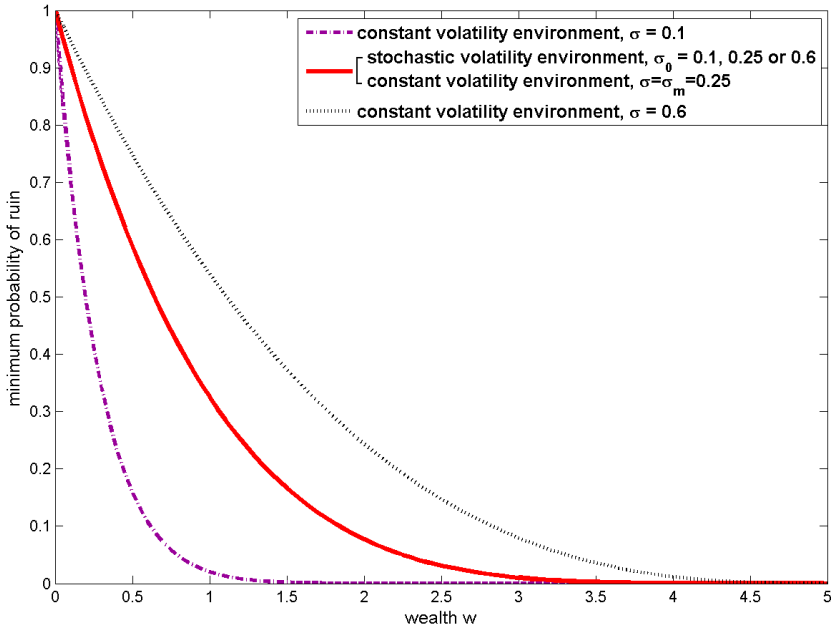


(b) Probability of ruin at $\sigma_0 = \sigma_m$=0.25

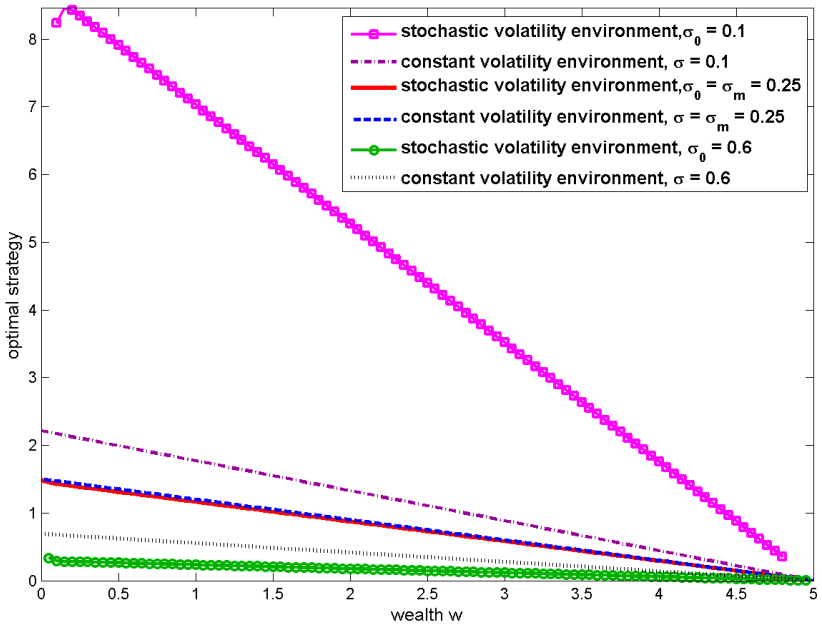


(c) Probability of ruin at σ_0=0.1

Figure 2.2: Variations of the minimum probability of ruin with respect to ρ. Speed of mean reversion= 0.5.

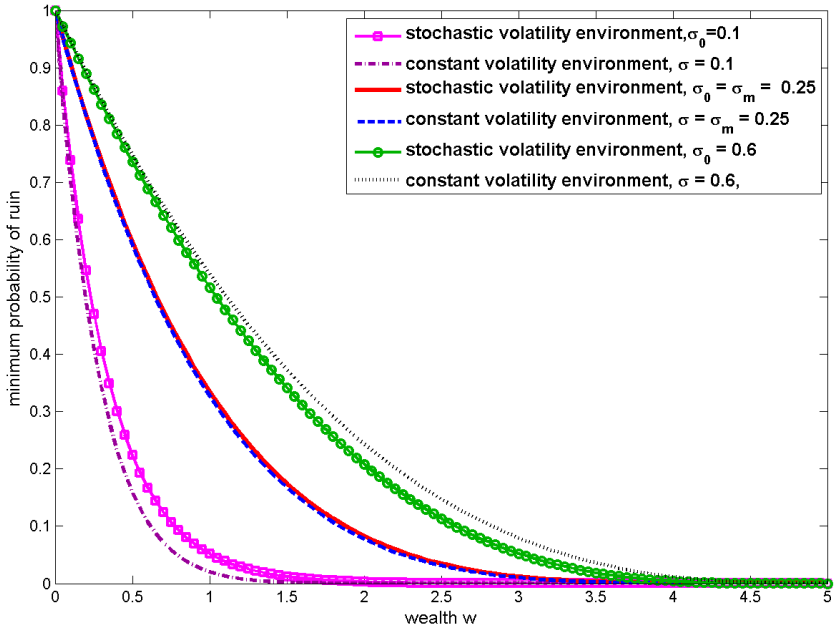


(a) Minimum probability of ruin

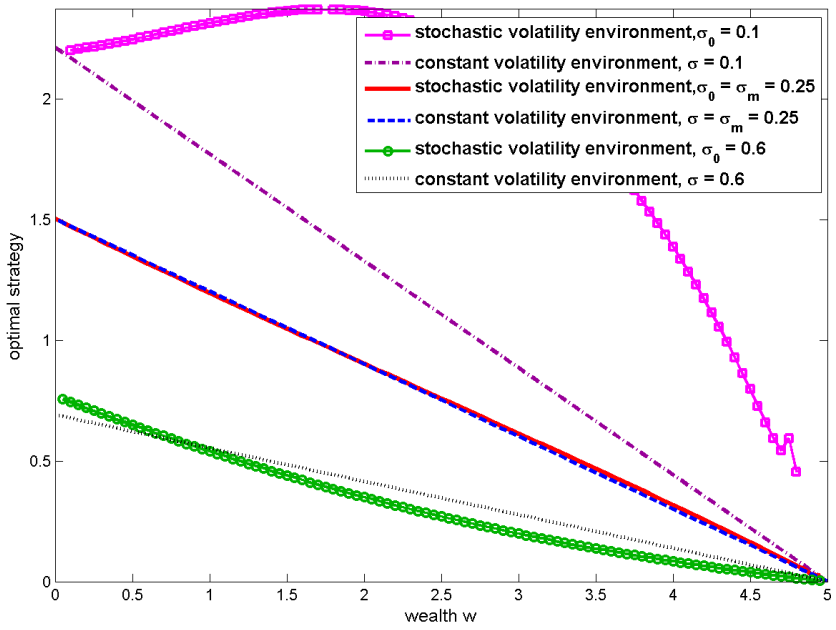


(b) Optimal investment strategy

Figure 2.3: Stochastic volatility versus constant volatility environment. Speed of mean reversion is 250.

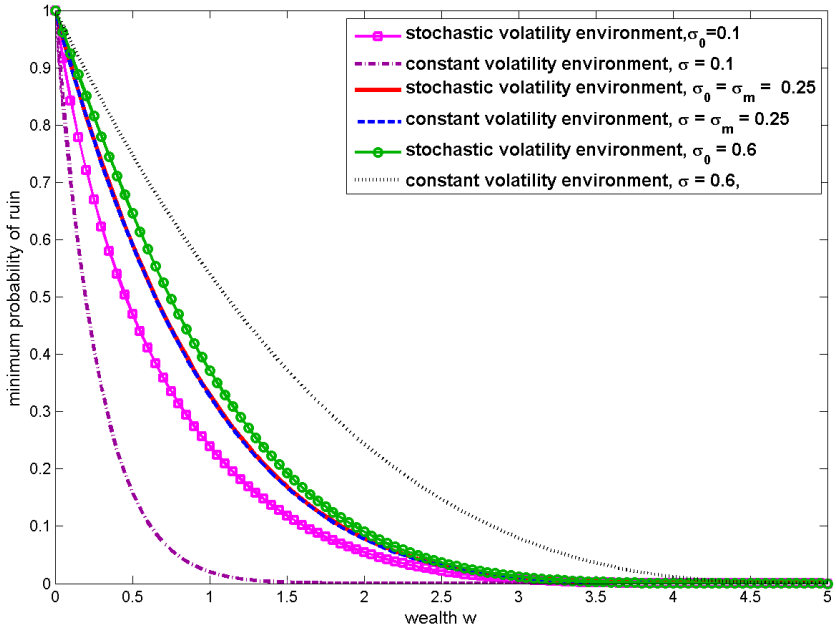


(a) Minimum probability of ruin

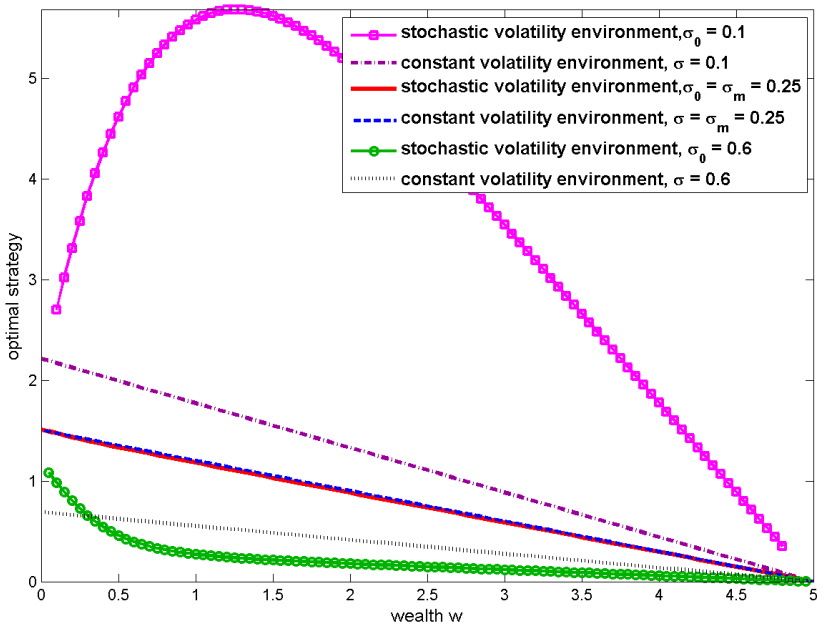


(b) Optimal investment strategy

Figure 2.4: Stochastic volatility versus constant volatility environment. Speed of mean reversion is 0.02.



(a) Minimum probability of ruin



(b) Optimal investment strategy

Figure 2.5: Stochastic volatility versus constant volatility environment. Speed of mean reversion is 0.5.

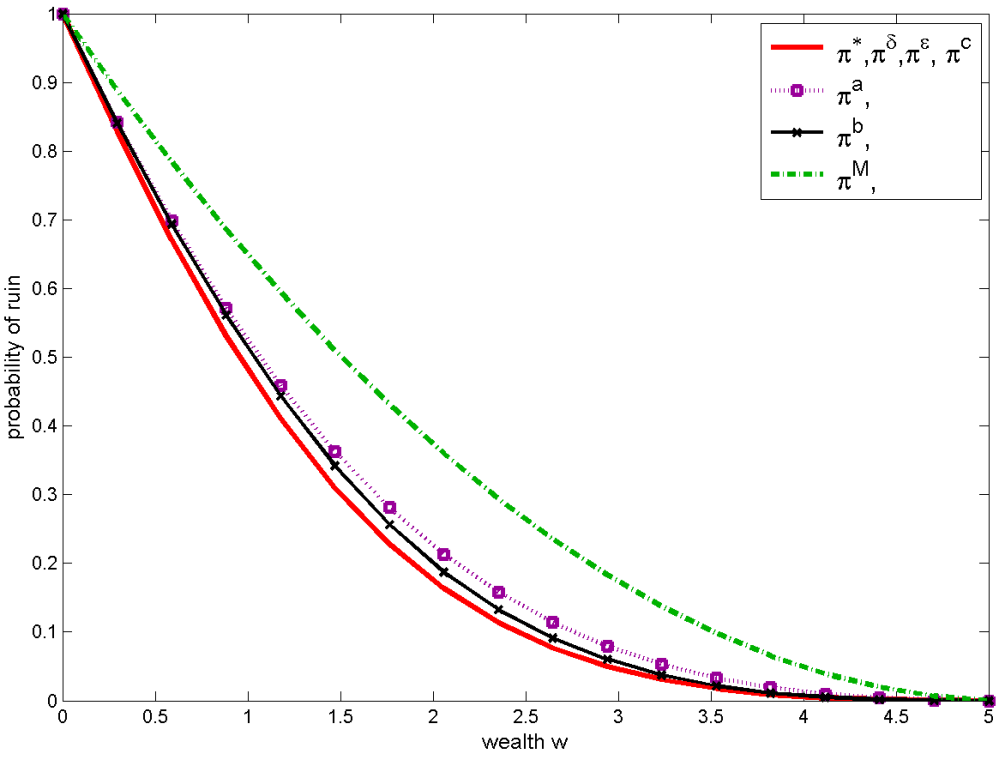


Figure 2.6: Performance of the investment strategies described in Observation 3 (of Section 6). Speed of mean reversion=0.2. Correlation $\rho = 0.5$. Initial volatility $\sigma_0 = 0.6$.

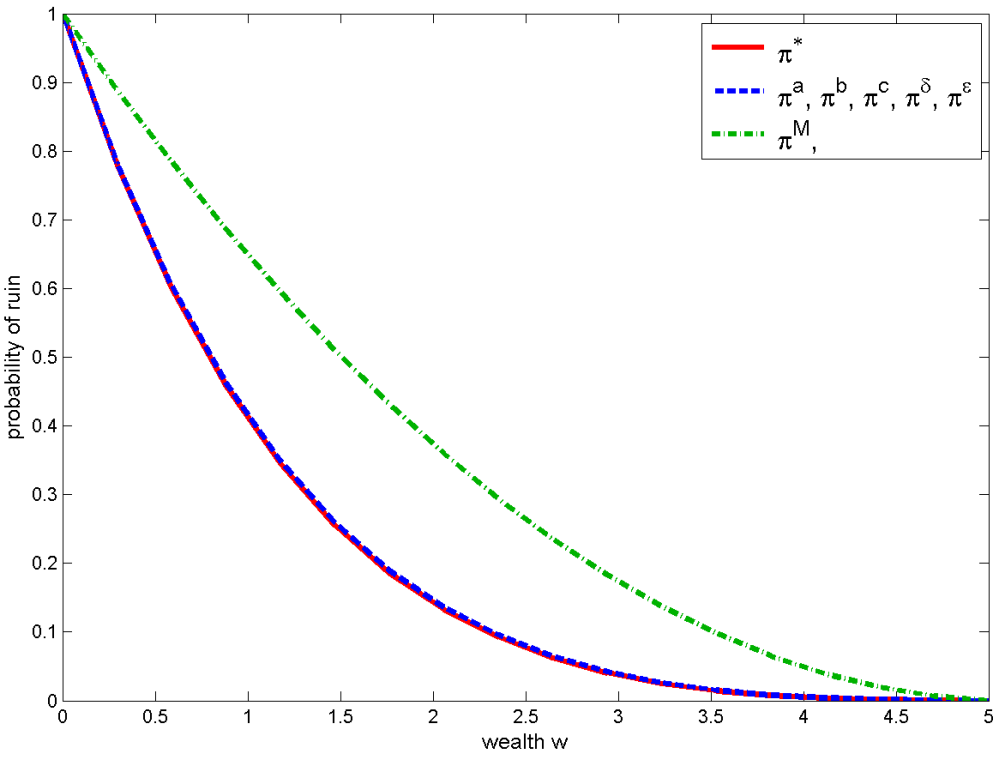


Figure 2.7: Performance of the investment strategies described in Observation 3 (of Section 6). Speed of mean reversion=0.2. Correlation $\rho = 0.5$. Initial volatility $\sigma_0 = \sigma_m = 0.25$.

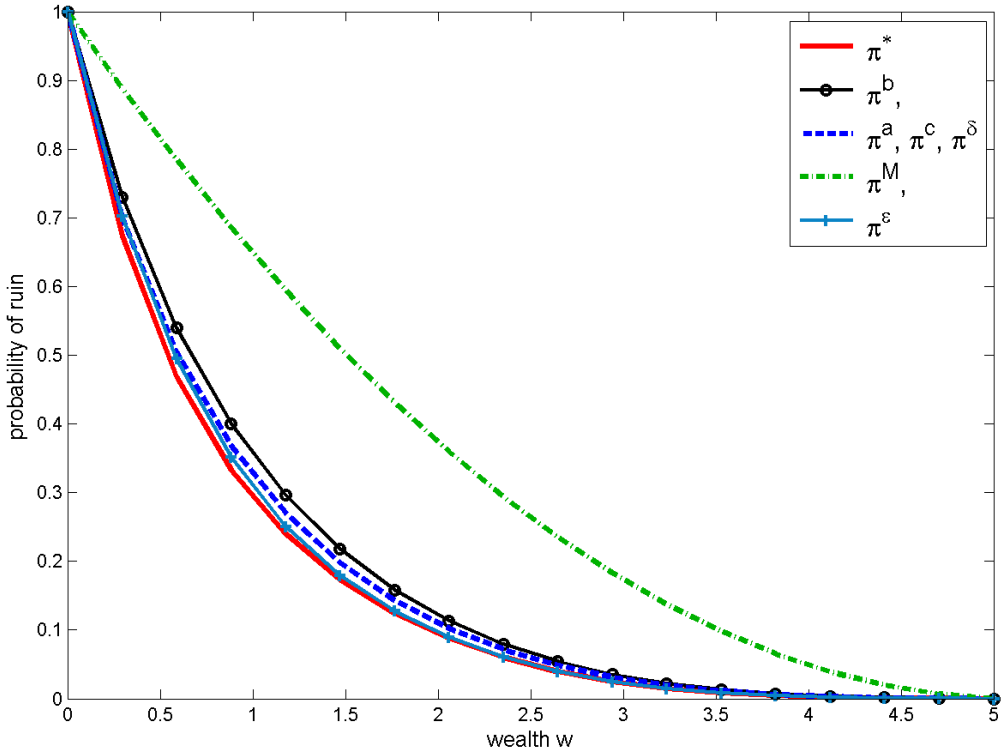


Figure 2.8: Performance of the investment strategies described in Observation 3 (of Section 6). Speed of mean reversion=0.2. Correlation $\rho = 0.5$. Initial volatility $\sigma_0 = 0.1$.

CHAPTER III

Exact Simulation of Heston Stochastic Volatility Model with Jumps

3.1 Introduction

The Heston stochastic volatility model is widely used in practice. Its popularity in equity, foreign exchange and interest rate markets rests on its computational tractability: Heston [61] showed that the value of a European call option can be obtained by transform inversion, facilitating quick calibration to market prices. However, the valuation of path-dependent and other securities with complex features is less tractable, and typically requires the application of Monte Carlo methods.

The Monte Carlo simulation of the Heston model is challenging. Standard discretization schemes are difficult to apply, largely because of the square-root diffusion dynamics of the volatility. These dynamics violate the conditions often used to guarantee convergence of a discretization scheme. Alfonsi [3], Andersen [4], Jaeckel and Kahl [63], Lord et al. [75], and Van Haastrecht and Pelsser [96] discuss and evaluate the performance of alternative discretization schemes. While relatively easy to implement, discretization methods introduce bias into the simulation estimator. The magnitude of the bias is hard to quantify. Therefore, it is difficult to obtain valid confidence intervals. Many time steps may be required to reduce the bias to an acceptable level, and a large number of simulation trials may be needed to verify that

the bias is sufficiently small. Finally, the optimal allocation of the computational budget between the number of time steps and the number of trials is difficult to specify in advance. In an innovative paper, Broadie and Kaya [27] develop an exact simulation scheme that addresses these issues. Glasserman and Kim [52] derive an explicit representation of the transitions of the Heston model and use it to circumvent the most time-consuming step in that scheme, namely the numerical inversion of the characteristic function of the integrated variance. Smith [94] approximates the characteristic function to improve computational efficiency. The cost of these improvements is a small bias.

In this chapter, we extend the Heston model to include state-dependent jumps, and develop a method for the exact simulation of this model. The price and the volatility are subject to common and correlated jumps. The jumps arrive with an intensity that may depend on time, price, volatility and jump counts. The jump size may depend on the price and volatility. The Monte Carlo algorithm extends a projection method developed by Giesecke et al. [49] for the exact simulation of a point process with an intensity driven by a one-dimensional jump-diffusion process. Filtering arguments, coupled with the analytical features of the square-root diffusion governing the volatility between jumps, facilitate the exact sampling of the Heston jump-diffusion in the filtration generated by the jumps. The samples from the exact distribution can then be used to generate an unbiased estimator of the price of a derivative security. We compare our method with the discretization scheme of Lord et al. [75], which appears to produce the smallest bias among several discretization schemes for the Heston model without jumps. We verify that the error of our simulation estimator converges faster. We also demonstrate that our scheme requires the smaller computational budget to achieve a given error.

The extension of Heston's stochastic volatility model to include common and correlated jumps in the price and volatility is empirically motivated. According to Eraker et al. [41], jumps in price and jumps in volatility serve complementary purposes. The former generate infrequently large movements, such as the 1987 crash. The latter generate sudden changes in the volatility and have a lasting impact on the distribution of price changes due to volatility persistence. The need to incorporate jumps in volatility in addition to jumps in return has been recognized by many authors, including Bakshi et al. [11], Duffie et al. [36], Eraker et al. [41], and others. Todorov and Tauchen [95] find evidence of common jumps in price and volatility as well as of component-specific jumps. Bates [13], Eraker [40], Pan [84] and others emphasize the state-dependence of jump arrivals. The Heston jump-diffusion model we propose can capture each of the aforementioned empirical features. The formulation can be further generalized to address a possible time-dependence of the coefficient functions.

Prior research has studied the exact simulation of jump-diffusion processes. Giesecke and Smelov [51] develop an exact acceptance/rejection scheme for a general one-dimensional jump-diffusion process with state-dependent drift, volatility, jump intensity and jump size distribution. Their method does not easily generalize to the two-dimensional Heston jump-diffusion. The exact method of Broadie and Kaya [27] extends to the case of jumps in the price and of contemporaneous jumps in the price and volatility, assuming that jumps arrive according to a Poisson process with constant intensity. Under this assumption, the jumps can be generated independently of the diffusion terms. The scope of our exact method includes state-dependent jump times and sizes. With state-dependence, the diffusion and jump terms cannot be simulated independently. We address this issue by proceeding sequentially. After

generating the jump times and sizes in their own filtration, we draw the value of the jump-diffusion at the horizon from its conditional distribution given the jump times and sizes. The conditional distribution is encoded in the conditional transform, which follows from point process filtering arguments. This method is the first example of an exact algorithm for a multi-dimensional jump-diffusion with correlated diffusion terms and state-dependent jump times and sizes.

The rest of this chapter is organized as follows. Section 3.2 introduces the Heston jump-diffusion model and explains a scheme for its discretization. Section 3.3 explains the exact simulation method. The algorithm for exact simulation is given in Section 3.4. Section 3.5 provides numerical results and Section 4.4.4 concludes. There is a technical Appendix.

3.2 The Heston Jump-diffusion

3.2.1 Model specification

Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying the usual conditions. Let S be a strictly positive process representing the stock price under the risk neutral measure $\mathbb{P}$, and let $Y = \log(S)$ be the log-price process. The variance process of the stock is denoted by V. The risk-neutral dynamics of the state process (Y, V) are given by

$$\begin{aligned} dY_t &= \left[r(t) - V_t/2 - A_t\right]dt + \sqrt{V_t}\left[\rho(t)dW_t^v + \sqrt{1-\rho^2(t)}dW_t^y\right] + dJ_t^y, \\ dV_t &= \kappa(t)\left[\theta(t) - V_t\right]dt + \sigma(t)\sqrt{V_t}dW_t^v + \delta dJ_t^v. \end{aligned} \tag{3.2.1}$$

Here, $r(t)$ is the risk-free interest rate and $\sqrt{V_t}$ is the volatility. The variance process V follows a mean-reverting square-root process with long-run mean $\theta(t)$, mean reversion speed $\kappa(t)$, and volatility $\sigma(t)$. The process (W^y, W^v) is a standard Brownian motion. The instantaneous correlation between log-price and variance, $\rho(t)$, takes

values in $(-1, 1)$. The parameter δ specifies the sensitivity to jumps for the variance process. The compensator process $A_t = A(t, Y_t, V_t)$ will be specified later by no-arbitrage condition in risk-neutral measure. J^y and J^v are pure jump processes whose jumps have state-dependent sizes and arrive with state-dependent intensity.

To be precise, denoting

$$\boldsymbol{\mu}(t,y,v) = \begin{pmatrix} r(t) - v/2 - A(t,y,v) \\ \kappa(t)[\theta(t) - v] \end{pmatrix}, \quad \Sigma(t,y,v) = \sqrt{v} \begin{pmatrix} \rho(t) & \sqrt{1-\rho^2(t)} \\ \sigma(t) & 0 \end{pmatrix}, \tag{3.2.2}$$

we suppose (Y, V, J^y, J^v) has an infinitesimal generator $\mathcal{D}$, defined for bounded C^2 function f, as

$$\begin{aligned}
\mathcal{D}f(t,y,v,j^y,j^v) &= f_t(t,y,v,j^y,j^v) \\
&+ \nabla f(t,y,v,j^y,j^v) \cdot \boldsymbol{\mu}(t,y,v) + \frac{1}{2}\mathrm{tr}[\nabla^2 f(t,y,v,j^y,j^v)\Sigma(t,y,v)\Sigma'(t,y,v)] \\
&+ \Lambda_1(t,y,v) \int_{\mathbf{R}} [f(t, y + \Delta^y(t,y,v;z), v, j^y + \Delta^y(t,y,v;z), j^v) - f(t,y,v,j^y,j^v)]\, d\nu^1(z) \\
&+ \Lambda_2(t,y,v,j^y,j^v) \int_{\mathbf{R}} [f(t, y, v + \delta\Delta^v(t,y,v;z), j^y, j^v + \Delta^v(t,y,v;z)) - f(t,y,v,j^y,j^v)]\, d\nu^2(z) \\
&+ \Lambda_3(t,y,v) \int_{\mathbf{R}^2} \big[f(t, y + \Delta^{y,c}(t,y,v;z), v + \delta\Delta^{v,c}(t,y,v;z), j^y + \Delta^{y,c}(t,y,v;z), \\
&\qquad j^v + \Delta^{v,c}(t,y,v;z)) - f(t,y,v,j^y,j^v)\big] d\nu^3(z).
\end{aligned} \tag{3.2.3}$$

where $\nabla f = (f_y, f_v)'$, $\nabla^2 f$ is the Hessian matrix of f with respect to (y, v), measurable functions $\Delta^y, \Delta^v, \Delta^{y,c}, \Delta^{v,c}$ specify the jump magnitude and ν^i represents the probability distribution of marker z.

Intuitively, this means that, conditional on the path of the state process, there are three types of jumps:

(i) jumps in only Y, which arrive with stochastic intensity $\lambda_t^1 = \Lambda_1(t, Y_t, V_t)$ and have state-dependent jump sizes determined by function Δ^y;

(ii) jumps in only V, which arrive with stochastic intensity $\lambda_t^2 = \Lambda_2(t, Y_t, V_t, J_t^y, J_t^v)$ and have state-dependent jump sizes determined by function Δ^v;

(iii) common jumps in both Y and V, which arrive with stochastic intensity $\lambda_t^3 = \Lambda_3(t, Y_t, V_t)$ and have state-dependent jump sizes determined by the pair of functions $(\Delta^{y,c}, \Delta^{v,c})$.

Assume

$$\mathbb{E}\left[\int_0^t \lambda_s^i ds\right] < +\infty. \quad \forall t \in [0, T] \tag{3.2.4}$$

The no-arbitrage condition in risk-neutral measure requires that

$$\begin{aligned} A_t = A(t, Y_t, V_t) = &\Lambda_1(t, Y_t, V_t) \int_R \left(e^{\Delta^y(Y_t, V_t; z)} - 1\right) d\nu^1(z) + \\ &\Lambda_3(t, Y_t, V_t) \int_{R^2} \left(e^{\Delta^{y,c}(t, Y_t, V_t; z)} - 1\right) d\nu^3(z). \end{aligned}$$

To ensure the existence and uniqueness of the solution to SDE (3.2.1), technical restrictions need to be imposed, as discussed by Ikeda and Watanabe [62], Protter [87], Situ [93], Athreya et al. [10], Platen and Bruti-Liberati [86], Kliemann et al. [67], and Ceci and Gerardi [31]. As shown in Appendix A of Ceci and Gerardi [31] (page 571 - 572), a unique strong solution exists under assumption (A) while a solution exists in strong sense and is weakly unique under assumption (B). In this chapter we adopt those assumptions from Ceci and Gerardi [31].

3.2.2 Discretization scheme

A discretization method can be used to approximate the process (Y, V) between jump times. Partition time into segments of equal length h and let $(\widehat{Y}, \widehat{V})$ denote the discretization of (Y, V). Lord et al. [75] find that the following scheme produces the smallest discretization bias among a set of alternative schemes for the discretization of the Heston model:

$$\begin{aligned}\widehat{Y}_{i+1} &= \widehat{Y}_i + \big[r(t_i) - \frac{1}{2}\widehat{V}_i^+ - A(t_i, \widehat{Y}_i, \widehat{V}_i^+)\big]h \\ &\quad + \sqrt{\widehat{V}_i^+}\big[\rho(t_i)\Delta W_i^v + \sqrt{1-\rho(t_i)^2}\Delta W_i^y\big], \qquad (3.2.5)\\ \widehat{V}_{i+1} &= \widehat{V}_i + \kappa(t_i)\big[\theta(t_i) - \widehat{V}_i^+\big]h + \sigma(t_i)\sqrt{\widehat{V}_i^+}\Delta W_i^v, \qquad (3.2.6)\end{aligned}$$

where $\widehat{V}_i^+ = \max(\widehat{V}_i, 0)$. The variables ΔW_i^y and ΔW_i^v are i.i.d. normal with mean 0 and variance h.

The jump times of (Y, V) can be generated using a time-scaling scheme based on a result of Meyer [79]. Let N^i be the counting process associated with type i jumps, with stochastic intensity λ^i. The process $N = N^1 + N^2 + N^3$ counting all jumps of (Y, V) has stochastic intensity $\lambda = \lambda^1 + \lambda^2 + \lambda^3$. Under a change of time defined by $\int_0^t \lambda_s ds$, N is a standard Poisson process. This implies that the jump times (T_n) of N can be constructed as

$$T_{n+1} = \inf\left\{t \geq T_n : \int_{T_n}^t \lambda_s ds \geq \mathcal{E}_n\right\}, \qquad (3.2.7)$$

where $(\mathcal{E}_n)$ is a sequence of i.i.d. standard exponential variables and $T_0 = 0$. The complete path of the cumulative intensities $H_t = \int_{T_n}^t \lambda_s ds$ is approximated by Riemann sum on the discrete-time grids as

$$\widehat{H}_m = \sum_{T_n \leq t_k \leq t_m} \widehat{\lambda}_k h, \quad \text{for } h = t_k - t_{k-1}, \qquad (3.2.8)$$

where $\widehat{\lambda}_k$ is the sample of the continuous-time intensity process λ_t at time $t_k = kh$ on the grid. The next jump time is then determined by

$$T_{n+1} = \min\left\{t_m : \widehat{H}_m \geq \mathcal{E}_n, \quad \mathcal{E}_n \sim \text{ Exponential}(1)\right\}$$

Note that $\widehat{H}_0 = \widehat{\lambda}_{T_n} h$ and that if the hitting occurs between the grid point t_m and t_{m+1}, then the left endpoint t_m is taken as the hitting time. This scheme for

integration and computation of the hitting time is found, in Giesecke et al. [50], to perform the best among others in terms of error convergence.

After a new jump time $T_{n+1} = t_m$ is generated, we apply thinning algorithm by generating a marker ξ_{n+1} which takes the value $i \in \{1, 2, 3\}$ with probability $p_i = \widehat{\lambda}^i_m / \widehat{\lambda}_m$. This marker specifies the jump type and determines whether this jump time is for only one of Y_t and V_t or for a common jump in both of them. The jump sizes ΔJ^y and ΔJ^v are then sampled from their specific distribution accordingly. That is, if $\xi_{n+1} = 1$, then $(\Delta J^y, \Delta J^v) = (\Delta^y(t_m, \widehat{Y}_m, \widehat{V}_m; z), 0)$ where $z \sim \nu^1$; if $\xi_{n+1} = 2$, then $(\Delta J^y, \Delta J^v) = (0, \Delta^v(t_m, \widehat{Y}_m, \widehat{V}_m; z)$ where $z \sim \nu^2$; else if $\xi_{n+1} = 3$, then $(\Delta J^y, \Delta J^v) = (\Delta^{y,c}(t_m, \widehat{Y}_m, \widehat{V}_m; z), \Delta^{v,c}(t_m, \widehat{Y}_m, \widehat{V}_m; z))$ where $z \sim \nu^3$.

The log-price and variance process are updated at the jump time t_m as

$$\begin{pmatrix} \widehat{Y}_m \\ \widehat{V}_m \end{pmatrix} = \begin{pmatrix} \widehat{Y}_m \\ \widehat{V}_m \end{pmatrix} + \begin{pmatrix} \Delta J^y \\ \delta \Delta J^v \end{pmatrix},$$

where $\widehat{Y}_m$ and $\widehat{V}_m$ on the right-hand side of above equation represent the simulated state processes right before the jump time according to diffusion component using (3.2.5) (3.2.6), and those on the left-hand side represent state processes updated at the jump time. Similarly, we can update the point processes $\widehat{J}^y_m, \widehat{J}^v_m$ and the intensities $\widehat{\lambda}^i_m$, $\widehat{\lambda}_m$.

In summary, we have the following algorithm to obtain a simulation path for $(\widehat{Y}, \widehat{V}), (\widehat{J}^y, \widehat{J}^v)$ given Y_0 and V_0 using Euler scheme.

Algorithm 3.2.1. Pseudo-code of the Euler scheme for the Heston model with stochastic intensity:

1: Initialize $(\widehat{Y}_0, \widehat{V}_0) \leftarrow (Y_0, V_0), (\widehat{J}^y_0, \widehat{J}^v_0) \leftarrow (0, 0)$, $n \leftarrow 0, T_n \leftarrow 0$.

2: Set $H \leftarrow \widehat{\lambda}_0 h$, and generate $\mathcal{E} \sim$ Exponential(1).

3: **for** $k = 1, \cdots, M$ **do**

4: set $t_k \leftarrow kh$;

5: simulate the (pre-jump) state process $\widehat{Y}_k, \widehat{V}_k$ given $\widehat{Y}_{k-1}$ and $\widehat{V}_{k-1}$ using (3.2.5) and (3.2.6);

6: compute the (pre-jump) intensity $\widehat{\lambda}_k = \Lambda(t_k, (\widehat{Y}_k, \widehat{V}_k), (\widehat{J}^y_k, \widehat{J}^v_k))$;

7: set the (pre-jump) point processes $(\widehat{J}^y_k, \widehat{J}^v_k) \leftarrow (\widehat{J}^y_{k-1}, \widehat{J}^v_{k-1})$;

8: $H \leftarrow H + \widehat{\lambda}_k h$;

9: **if** $H \geq \mathcal{E}$ **then**

10: set the jump time $T_{n+1} \leftarrow t_k$;

11: compute the intensities $\widehat{\lambda}^i_k = \Lambda_i(t_m, (\widehat{Y}_m, \widehat{V}_m), (\widehat{J}^y_m, \widehat{J}^v_m))$ for $i = 1, 2, 3$;

12: generate a marker ξ taking value $i \in \{1, 2, 3\}$ with probability $\widehat{\lambda}^i_k / \widehat{\lambda}_k$, and generate the jump sizes ΔJ^y and ΔJ^v accordingly, similar as in Algorithm 3.4.3;

13: update the (post-jump) state process as $\widehat{Y}_k \leftarrow \widehat{Y}_k + \Delta J^y$ and $\widehat{V}_k \leftarrow \widehat{V}_k + \delta \Delta J^v$.

14: update the (post-jump) point process as $\widehat{J}^y_k \leftarrow \widehat{J}^y_k + \Delta J^y$ and $\widehat{J}^v_k \leftarrow \widehat{J}^v_k + \Delta J^v$.

15: update the (post-jump) intensity as $\widehat{\lambda}_k = \Lambda(t_k, (\widehat{Y}_k, \widehat{V}_k), (\widehat{J}^y_k, \widehat{J}^v_k))$

16: reset $H \leftarrow \widehat{\lambda}_k h$ and generate a new $\mathcal{E} \sim$ Exponential(1)

17: $n \leftarrow n + 1$

18: **end if**

19: **end for**

Let $\hat{Y}^M_T$ and $\hat{V}^M_T$ be the simulated value of the log-price and variance of stock price using M time-steps. The price of a financial derivative $\mathbb{E}[f(S_T, V_T)]$ can then

be estimated by Monte Carlo simulation as

$$C^{N,M} = \frac{1}{N}\sum_{k=1}^{N} f(\exp(\hat{Y}_T^M), \hat{V}_T^M),$$

where N is the number of sample paths. This estimator converges to the derivative's price as the number of sample paths N and number of time steps M get larger.

There are multiple layers of bias in this Monte-Carlo estimation with Euler discretization. First, there is a discretization error or bias caused by using discrete-time approximation (3.2.5),(3.2.6) in the simulation of continuous-time state processes between jumps. For a pure diffusion process, under some regularity conditions (see Kloeden and Platen [68], Bally and Talay [12]), the discretization error decreases at the rate $O(1/M)$ for Euler discretization. For the SDE in the Heston model, however, the conditions for first-order convergence do not hold, so the actual convergence may be slower. When state-dependent jumps are involved, the convergence rate becomes more complicated to determine and is largely unknown. Second, since the jump times are determined as the first hitting time of the integrated intensity to an exponential level, there is an approximation error or bias by approximating the paths of cumulative intensities in (3.2.7) on discrete-time grid, and by approximating the integral with Riemann sum. Other integration schemes, for example, the trapezoidal rule, could be used instead. The improvements of the basic integration scheme, however, may not give a better error convergence. In addition, even if one can eliminate the discretization bias mentioned previously and simulate the state processes at grid points exactly, the bias in hitting time would remain.

Because of these multiple layers of errors or bias in the Monte-Carlo estimation with Euler discretization, it is difficult to quantify the magnitude of the error in the estimator of financial derivative's price, and hence it is hard to obtain valid confidence intervals. In addition, it is difficult to specify the optimal allocation of

computational resources between number of paths N and number of steps M.

3.3 Exact Simulation

This section explains an algorithm for the exact simulation of the Heston jump diffusion (3.2.1). The algorithm exploits the *projection method* developed by Giesecke et al. [49] for the exact simulation of point processes with stochastic intensities. We project the point process (J^y, J^v) representing the jumps of (Y, V) onto its own filtration, and generate the event times based on the arrival intensity in this sub-filtration. The sub-filtration intensity is deterministic between arrivals, and this property facilitates the application of exact schemes for the sequential generation of event times. Given a path of (J^y, J^v) over some period $[0, T]$, we then draw the value of (Y_T, V_T) from the conditional transform given the point process paths. The conditional transform is computed using point process filtering arguments. This approach applies to a wide range of multi-variate jump-diffusion processes with state-dependent jump intensities.

We write T_n as the nth arrival time of the global counting process N. The counting processes N^i would be extracted later from the global counting process N by thinning algorithm.

3.3.1 Intensity projection

The idea of our unbiased exact simulation, as mentioned before, is to project the jump processes onto their own right-continuous and complete filtration $\mathcal{G}_t = \sigma(J^y_s, J^v_s; s \leq t)$. After the projection, we have the inter-arrival intensity

$$h_i(t) = \mathbb{E}[\lambda^i_t | \mathcal{G}_t] = \sum_n h_{i,n}(t) 1_{\{N(t)=n\}},$$

and

$$h(t) = \mathbb{E}[\lambda_t | \mathcal{G}_t] = h_1(t) + h_2(t) + h_3(t).$$

They are deterministic between events and jump at event times.

To calculate the projected intensity, we need the conditional distribution $\pi_t(dy, dv)$ of the state processes (Y_t, V_t) given $\mathcal{G}_t$. The distribution is implied by the Laplace transform

$$M_t(z_1, z_2) = \mathbb{E}[e^{-z_1 Y_t - z_2 V_t} | \mathcal{G}_t] = \int_{\mathbf{R}\times\mathbf{R}^+} e^{-z_1 y - z_2 v} \pi_t(dy, dv).$$

We will refer to $M_t(z_1, z_2)$ as the filter.

The filtering results for point processes (see Kliemann et al. [67], Ceci and Gerardi [31], Frey et al. [46]) imply that $M_t(z_1, z_2)$ is the solution to the Kushner-Stratonovich equation. The assumption (A) or (B) in Section 2.1 guarantee that the solution to the Kushner-Stratonovich equation is unique. The filtering equation splits to an equation for $t \in [T_n, T_{n+1})$ between arrivals and an update at arrival T_{n+1}.

3.3.2 Filtering between arrivals of jumps

To calculate the filter, we define an auxiliary process $\mathbf{X} = (X^y, X^v)'$ as a unique solution to the stochastic differential equation

$$d\mathbf{X}_t = \mathbf{x} + \int_0^t \boldsymbol{\mu}(s, \mathbf{X}_s) ds + \int_0^t \Sigma(s, \mathbf{X}_s) d\mathbf{W}_s, \tag{3.3.1}$$

where the drift function $\boldsymbol{\mu}$ and volatility matrix Σ are the same as in (3.2.2). For $s \le t$, $\mathbf{x} = (x_1, x_2) \in \mathbf{R}^2, \mathbf{J} = (j^y, j^v) \in \mathbf{R}^2$, let [1]

$$\varphi(s, t, \mathbf{x}, \mathbf{J}; z_1, z_2) = \mathbb{E}\left[\exp\left\{-\int_s^t \Lambda(u, \mathbf{X}_u, \mathbf{J}) du\right\} e^{-z_1 X_t^y - z_2 X_t^v} \Big| \mathbf{X}_s = \mathbf{x}\right], \tag{3.3.2}$$

Theorem 4.1 in Kliemann et al. [67] implies that, for $t \in [T_n, T_{n+1})$,

$$M_t(z_1, z_2) = \frac{\rho_t(z_1, z_2)}{\rho_t(0, 0)}, \tag{3.3.3}$$

where

$$\rho_t(z_1, z_2) = \mathbb{E}[\varphi(T_n, t, (Y_{T_n}, V_{T_n}), (J_{T_n}^y, J_{T_n}^v); (z_1, z_2)) | \mathcal{G}_t].$$

[1] We use bold font for vectors, that is, $\mathbf{x} = (x_1, x_2)$, $\mathbf{z} = (z_1, z_2)$, $\mathbf{b} = (b_1, b_2)$, $\mathbf{l} = (l_1, l_2)$, $\mathbf{J}_t = (J_t^y, J_t^v)$, and so on.

3.3.3 Filtering at arrival of jumps

Consider the jumps arrival time T_{n+1}. Let $\xi_{n+1} \in \{1, 2, 3\}$ be the variable that classifies the types of the shocks at time T_{n+1}. From the filtering result (Kliemann et al. [67], Ceci and Gerardi [31], Frey et al. [46]), we have

$$\begin{aligned} M_{T_{n+1}}(z_1, z_2) &= \frac{\mathbb{E}[\lambda_{T_{n+1}^-}^{(\xi_{n+1})} \exp\left(-z_1 Y_{T_{n+1}} - z_2 V_{T_{n+1}}\right) | \mathcal{G}_{T_{n+1}^-}]}{\mathbb{E}[\lambda_{T_{n+1}^-}^{(\xi_{n+1})} | \mathcal{G}_{T_{n+1}^-}]} \\ &= e^{-z_1 \Delta J_{n+1}^y - z_2 \delta \Delta J_{n+1}^v} \frac{\mathbb{E}[\lambda_{T_{n+1}^-}^{(\xi_{n+1})} \exp(-z_1 Y_{T_{n+1}^-} - z_2 V_{T_{n+1}^-}) | \mathcal{G}_{T_{n+1}^-}]}{\mathbb{E}[\lambda_{T_{n+1}^-}^{(\xi_{n+1})} | \mathcal{G}_{T_{n+1}^-}]}, \end{aligned} \tag{3.3.4}$$

where we assume the jump sizes $(\Delta J_{n+1}^y, \Delta J_{n+1}^v)$ are measurable with respect to $\mathcal{G}_{T_{n+1}^-}$.

3.3.4 Computing the filter

Concluding the above sub-sections, we see that an important step is to obtain a representation of φ in (3.3.2), which will then lead to a solution of the filtering equations.

By Ito's formula, φ satisfies the following partial differential equation,

$$\begin{aligned} &\frac{\partial \varphi}{\partial s} + \big(r(s) - A(s, y, v) - \frac{1}{2} v\big) \frac{\partial \varphi}{\partial y} + \kappa(s) \big(\theta(s) - v\big) \frac{\partial \varphi}{\partial v} \\ &\quad + \frac{1}{2} v \left(\frac{\partial^2 \varphi}{\partial^2 y} + \sigma^2(s) \frac{\partial^2 \varphi}{\partial v^2} + 2\rho(s)\sigma^2(s) \frac{\partial^2 \varphi}{\partial y \partial v} \right) = \Lambda(s, y, v, \mathbf{J}) \varphi \end{aligned} \tag{3.3.5}$$

with terminal condition $\varphi(s, y, v, \mathbf{J}; z_1, z_2)|_{s=t} = \exp(-z_1 y - z_2 v)$ for fixed z_1, z_2 and t.

The representation (3.3.2) of φ is also closely related to the term-structure of bond price in interest rate theory. In particular, closed-form formulas are developed, for

example, in Duffie et al. [36] for affine setting, in Leippold and Wu [71] for quadratic setting.

If no closed-form for φ is available, some numerical recursive algorithm might be used to compute the filter. For example, Ceci and Gerardi [31] developed approximations to the filter based on Markov chain approximation method, and Frey and Runggaldier [47] developed approximations by particle filtering method.

In the next sub-section, we illustrate the affine case, where the explicit representation is available for φ and hence for the filter.

3.3.5 Filtering in affine models

If the drift vector $\boldsymbol{\mu}$, volatility matrix Σ, and the jumps intensities have affine structures with respect to the state processes (Y, V), we can obtain semi-analytical recursive formulas for projected intensity and the filter $M_t(z_1, z_2)$. We assume the intensities are affine in (Y, V) as

$$\begin{aligned}
&\lambda_t^i = \Lambda_i(t, (Y_t, V_t)) = \Lambda_{i,0}(t) + \Lambda_{i,1}(t)Y_t + \Lambda_{i,2}(t)V_t, \quad i = 1, 3, \\
&\lambda_t^2 = \Lambda_2(t, (Y_t, V_t), (J_t^y, J_t^v)) = \Lambda_{2,0}(t, (J_t^y, J_t^v)) + \Lambda_{2,1}(t, (J_t^y, J_t^v))Y_t + \Lambda_{2,2}(t, (J_t^y, J_t^v))V_t,
\end{aligned} \tag{3.3.6}$$

therefore the global intensity $\lambda_t = \lambda_t^1 + \lambda_t^2 + \lambda_t^3$ is also affine as

$$\lambda_t = \Lambda(t, (Y_t, V_t), ((J_t^y, J_t^v)) = \Lambda_{0,0}(t, (J_t^y, J_t^v)) + \Lambda_{0,1}(t, (J_t^y, J_t^v))Y_t + \Lambda_{0,2}(t, (J_t^y, J_t^v))V_t. \tag{3.3.7}$$

Also assume the drift term $\boldsymbol{\mu}(t, y, v)$ is affine in (Y, V). Notice that the drift of V is already affine as $\kappa(t)(\theta(t) - V_t)$, we only need to assume the drift of Y be affine as $\mu_0(t) - \mu_1(t)Y_t - \mu_2(t)V_t$.

Since the auxiliary process $\mathbf{X}$ is now a two-dimensional affine diffusion, thanks to

Duffie et al. [36], we have the formula for φ as

$$\varphi(s,t,\mathbf{x},\mathbf{J};\mathbf{z}) = \exp\left(a(s,t,\mathbf{J},\mathbf{z}) - b(s,t,\mathbf{J},\mathbf{z})\cdot\mathbf{x}\right),$$

where functions $\mathbf{b} = (b_1, b_2)$ and a satisfy the ODE

$$\begin{aligned}
\partial_s b_1(s,t,\mathbf{J},\mathbf{z}) &= -\Lambda_{0,1}(s,\mathbf{J}) + \mu_1(s)b_1,\\
\partial_s b_2(s,t,\mathbf{J},\mathbf{z}) &= -\Lambda_{0,2}(s,\mathbf{J}) + \mu_2(s)b_1 + \kappa b_2 + \frac{1}{2}(b_1^2 + 2\rho\sigma b_1 b_2 + \sigma^2 b_2^2),\\
\partial_s a(s,t,\mathbf{J},\mathbf{z}) &= \Lambda_{0,0}(s,\mathbf{J}) + \mu_0(s)b_1 + \kappa\theta b_2,
\end{aligned} \tag{3.3.8}$$

with terminal condition $b_1|_{s=t} = z_1, b_2|_{s=t} = z_2, a|_{s=t} = 0$.

Remark 3.3.1. Note that Duffie et al. [36] gives a closed-form formula for the Laplace transform of the distribution of state process when the intensity is affine in both the state processes and the point processes. Our exact simulation method can handle the case where intensities are not affine in the point processes (J^y, J^v), and hence goes beyond Duffie et al. [36].

3.4 Algorithm for exact simulation

Based on the above results, we can now write down an algorithm to recursively compute the filter and to exact simulate the Heston jump diffusion when the intensities are affine in state processes.

Algorithm 3.4.1. Algorithm for exact jump simulation

(i) Initialization: set $T_0 = 0$, compute $M_0(z_1, z_2) = \mathbb{E}[\exp(-z_1 Y_0 - z_2 V_0)]$;

(ii) Computing intensity projection: for $n = 1, 2, ..,$ for $t \in [T_n, T_{n+1})$ compute

$$\begin{aligned}
&h_n^i(t) = \mathbb{E}[\lambda_t^i|\mathcal{G}_t]\\
&= \Lambda_{i,0}(T_n, \mathbf{J}_{T_n}) - \Lambda_{i,1}(T_n, \mathbf{J}_{T_n})\partial_{z_1} M_t(z_1,z_2)\big|_{z_1=z_2=0} - \Lambda_{i,2}(T_n, \mathbf{J}_{T_n})\partial_{z_2} M_t(z_1,z_2)\big|_{z_1=z_2=0};
\end{aligned} \tag{3.4.1}$$

(iii) Simulating the jump time: generate the next jump times T_{n+1} according to the intensity $h_n(t) = h_n^1(t) + h_n^2(t) + h_n^3(t)$ (see Algorithm 3.4.2);

(iv) Thinning: classify the jump type ξ_{n+1} which takes value of $1, 2, 3$ with probability p_1, p_2, p_3 respectively, and generate sample jumps sizes $(\Delta J^y_{n+1}, \Delta J^v_{n+1})$ (see Algorithm 3.4.3);

(v) Update the filter recursively:

For $t \in [T_n, T_{n+1})$,

$$M_t(z_1, z_2) = \frac{\exp(a(T_n, t, \mathbf{J}_{T_n}, \mathbf{z}))M_{T_n}(\mathbf{b}(T_n, t, \mathbf{J}_{T_n}, \mathbf{z}))}{\exp(a(T_n, t, \mathbf{J}_{T_n}, \mathbf{0}))M_{T_n}(\mathbf{b}(T_n, t, \mathbf{J}_{T_n}, \mathbf{0}))},$$

and at jump time T_{n+1},

$$M_{T_{n+1}}(z_1, z_2) = A_{n+1}(z_1, z_2) \cdot \frac{\widetilde{\Lambda}_{n+1,0} M_{T^-_{n+1}}(z_1, z_2) - \widetilde{\Lambda}_{n+1,1}\partial_{z_1} M_{T^-_{n+1}}(z_1, z_2) - \widetilde{\Lambda}_{n+1,2}\partial_{z_2} M_{T^-_{n+1}}(z_1, z_2)}{[\widetilde{\Lambda}_{n+1,0} - \widetilde{\Lambda}_{n+1,1}\partial_{z_1} M_{T^-_{n+1}}(z_1, z_2) - \widetilde{\Lambda}_{n+1,2}\partial_{z_2} M_{T^-_{n+1}}(z_1, z_2)]|_{z_1=0, z_2=0}},$$

where

$$\widetilde{\Lambda}_{n+1,k} = \Lambda_{\xi_{n+1},k}(T^-_{n+1}, \mathbf{J}_{T_n}), k = 0, 1, 2,$$

$$A_{n+1}(z_1, z_2) = \exp(-z_1 \Delta J^y_{n+1} - z_2 \delta \Delta J^v_{n+1}),$$

$$M_{T^-_{n+1}}(z_1, z_2) = \lim_{t \to T^-_{n+1}} M_t(z_1, z_2).$$

(vi) Repeat the above steps, cursively compute $h_n(t)$ and $M_t(z_1, z_2)$, until we reach the time horizon T.

Algorithm 3.4.2 (Sub-routine used in Algorithm 3.4.1)**.** Algorithm for generating jump times T_{n+1} given $h_n(t)$:

(i) Set $t = T_n$;

(ii) Find bound B^n_t and C^n_t such that for $s \in [0, C^n_t]$, $h_n(t + s) \leq B^n_t$;

(iii) Generate candidate jump time $\hat{t} = t + \mathcal{E}$, where $\mathcal{E} \sim$ Exponential(B^n_t),

- If $\mathcal{E} > C^n_t$, set $t = \hat{t}$ and repeat step 2;

- Else, calculate the ratio $q = h_n(t+\mathcal{E})/B_t^n$, draw a standard uniform random variable $\mathcal{U}$,
  - If $\mathcal{U} \leq q$, accept the candidate, set $T_{n+1} = \hat{t}$, stop;
  - Else, reject the candidate, set $t = \hat{t}$, and repeat step 2.

Algorithm 3.4.3 (Sub-routine used in Algorithm 3.4.1)**.** Algorithm for classifying jump type and generate jump sizes:

(i) Calculate the ratios $p_i = \frac{h_n^i(T_{n+1})}{h_n(T_{n+1})}$, and generate a uniform random variable $\mathcal{V}$,

- If $\mathcal{V} < p_1$, then $\xi_{n+1} = 1$;
- Else if $\mathcal{V} < p_1 + p_2$, then $\xi_{n+1} = 2$;
- Else $\xi_{n+1} = 3$.

(ii) Classify the type of shock according to the result of jump type ξ_{n+1} and simulate jump sizes:

- If $\xi_{n+1} = 1$, only Y_t jumps; we set $\Delta J^v = 0$ and generate jump sizes $\Delta J^y = \Delta^y(t, Y_{T_{n+1}^-}, V_{T_{n+1}^-}; z)$ where z follows distribution $\nu^1(z)$;
- If $\xi_{n+1} = 2$, only V_t jumps; we set $\Delta J^y = 0$ and generate jump sizes $\Delta J^v = \Delta^v(t, Y_{T_{n+1}^-}, V_{T_{n+1}^-}; z)$ where $z \in \mathbf{R}$ is sampled from distribution $\nu^2(z)$;
- If $\xi_{n+1} = 3$, both Y_t and V_t jump; we generate jump sizes $\Delta J^y = \Delta^{y,c}(t, Y_{T_{n+1}^-}, V_{T_{n+1}^-}; z)$ and $\Delta J^y = \Delta^{y,c}(t, Y_{T_{n+1}^-}, V_{T_{n+1}^-}; z)$ where $z \in \mathbf{R}^2$ is sampled from the joint distribution $\nu^3(z)$,

Remark 3.4.4. Under the following cases, our exact simulation method degenerates to a conditioning transform method. The algorithm 3.4.1 still works, but is not a "projection method" anymore.

(i) $\lambda_t^i \equiv 0$. That is, jump occurs in neither the log-price process nor the variance process, and the model degenerates to the stochastic volatility (SV) model. In this cases, steps (2)-(5) is unnecessary since there are no jumps, and the filter at terminal time horizon is simply

$$M_T(z_1, z_2) = \frac{\exp(a(0, T, \mathbf{0}, \mathbf{z}))M_{T_n}(\mathbf{b}(0, T, \mathbf{0}, \mathbf{z}))}{\exp(a(0, T, \mathbf{0}, \mathbf{0}))M_{T_n}(\mathbf{b}(0, T, \mathbf{0}, \mathbf{0}))}.$$

(ii) λ_t^i are all deterministic functions with at least one of them non-zero. In this case, the projection step (2) is unnecessary, and the filter is simply updated at jump times in step (5) as

$$M_{T_{n+1}}(z_1, z_2) = A_{n+1}(z_1, z_2)\frac{\exp(a(T_n, t, \mathbf{J}_{T_n}, \mathbf{z}))M_{T_n}(\mathbf{b}(T_n, t, \mathbf{J}_{T_n}, \mathbf{z}))}{\exp(a(T_n, t, \mathbf{J}_{T_n}, \mathbf{0}))M_{T_n}(\mathbf{b}(T_n, t, \mathbf{J}_{T_n}, \mathbf{0}))}.$$

In this chapter, we do not consider these two cases and assume that at least one of the intensities λ_t^i is non-zero and stochastic. We also rule out the case where the jump sizes are simultaneously zero almost surely.

3.4.1 Exact method for a certain class of pay-off functions

We consider the price of a general financial derivative $\mathbb{E}[f(Y_T, V_T, J_T^y, J_T^v)]$ where the final payoff function takes the form $f(y, v, j^y, j^v) = \sum_{p,q,m,n} e^{py}e^{qv}y^m v^n g(j^y, j^v)$, where p, q are real numbers, m, n are integers and $g(\cdot)$ is some measurable function. In this case, the price of the derivative can be easily obtained using the filter calculated in Algorithm 3.4.1. To see this, notice that, for each term in the component of such function,

$$\begin{aligned}\mathbb{E}[S_T^p e^{qV_T} Y_T^m V_T^n g(J_T^y, J_T^v)] &= \mathbb{E}\big[g(J_T^y, J_T^v)\mathbb{E}[S_T^p e^{qV_T} Y_T^m V_T^n | \mathcal{G}_T]\big] \\ &= \mathbb{E}[g(J_T^y, J_T^v)(-1)^{m+n}\partial_{z_1}^{(m)}\partial_{z_2}^{(n)} M_T(z_1, z_2)|_{z_1=-p, z_2=-q}],\end{aligned} \tag{3.4.2}$$

where the derivatives $\partial_{z_1}^{(m)}\partial_{z_2}^{(n)}M_T(z_1,z_2)$ can be calculated from the filter $M_T(z_1,z_2)$.

For the kth path, we estimate the values of the conditional price as $C^{(k)} = \mathbb{E}[f(S_T,V_T,\mathbf{J}_T)|\mathcal{G}_T]$, then the Monte-Carlo estimator for the price of the financial derivative is given by

$$C^N = \frac{1}{N}\sum_{k=1}^{N} C^{(k)}. \tag{3.4.3}$$

Example 3.4.5 (Futures)**.** Futures on a stock have (un-discounted) price

$$F = \mathbb{E}[S_T].$$

Note that $\mathbb{E}[S_T|\mathcal{G}_T] = M_T(-1,0)$, hence the Monte-Carlo price of a future is

$$F^N = \frac{1}{N}\sum_{k=1}^{n} M_T^{(k)}(-1,0),$$

where the subscript (k) denotes estimator of the kth path.

Example 3.4.6 (Variance swap)**.** Consider the continuous strike variance swap rate, which is computed as the risk-neutral expectation of realized variance (see Carr and Wu [30]) as

$$C = \mathbb{E}[RV], \quad RV = \frac{1}{T}\int_0^T V_u du + \frac{1}{T}\sum_{n=1}^{N_T^1+N_T^3}\left(\Delta J_n^y\right)^2.$$

Define $I_t = \int_0^t V_s ds$, that is, $dI_t = V(t)dt$, and define

$$M_t(z_1,z_2,z_3) = \mathbb{E}[\exp(-z_1Y_t - z_2V_t - z_3I_t)|\mathcal{G}_t].$$

Notice the drift and volatility vector of (Y_t,V_t,I_t) are affine in the state processes if (Y_t,V_t) is so. The filtering algorithm can be easily extended to affine setting with three state processes. For the kth path,

$$C^{(k)} = \mathbb{E}[RV|\mathcal{G}_T] = -\frac{1}{T}\,\partial_{z_3}M_T(0,0,z_3)|_{z_3=0} + \frac{1}{T}\sum_{n=1}^{N_T^1+N_T^3}\left(\Delta J_n^y\right)^2,$$

hence the Monte-Carlo estimates of the variance swap rate can be easily obtained by (3.4.3).

3.4.2 Method I: FFT method for other pay-offs

The conditional distribution of (Y_t, V_t) given $\mathcal{G}_T$, denoted by $\pi_t(dy, dv)$, can be obtained by performing the inverse Laplace transform on $M_t(z_1, z_2)$. Exact samples of the state process can then be drawn from this distribution using the inverse method.

The characteristic function of log-price Y_T given $\mathcal{G}_T$ is given by $\phi(u) = \mathbb{E}[e^{iuY_T}|\mathcal{G}_T] = M_T(-iu, 0)$. Fourier inversion techniques, as used in Scott [92] and Broadie and Kaya [27], can be used to invert the characteristic function to sample for S_T conditional on $\mathcal{G}_T$. Let Y be a random variable with same distribution as Y_T given $\mathcal{G}_T$, then its cumulative distribution function can be written as

$$F(x) = \mathbb{P}(Y \leq x) = \frac{2}{\pi} \int_0^\infty \frac{sin(ux)}{u} \Re(M_T(-iu, 0))du, \tag{3.4.4}$$

where $\Re(c)$ denotes the real part of a complex number c.

Trapezoidal rule is then applied to compute it numerically as

$$F(x) = \frac{hx}{\pi} + \frac{2}{\pi} \sum_{j=1}^{L} \frac{sin(hjx)}{j} \Re(M_T(-ihj, 0)), \tag{3.4.5}$$

the number L is chosen so that

$$\frac{|M_T(-ihL, 0)|}{L} \leq \frac{\pi\epsilon}{2},$$

where ϵ is the desired truncation error.

The value of Y can then be simulated using inverse transform method as $Y = F^{-1}(\mathcal{U})$, where $\mathcal{U}$ is a uniform random number. Let $Y^{(k)}$ be the simulated value of Y for the kth path. Then the conditional price of a derivative with payoff $f(S_T)$ is given by $C^{(k)} = f(\exp(Y^{(k)}))$, and the Monte-Carlo price C^N is obtained by (3.4.3).

Alternatively, when we have the transform of the price of the financial derivative directly, (for example, the options), an easier and faster technique can be used to invert functionals of the characteristic function for only once on each path. In

particular, Carr and Madan [28] developed a FFT method for the price of a vanilla call option. Similar technique can be used here to calculate conditional call price $\mathbb{E}\big[(S_T-K)^+|\mathcal{G}_T\big]$ and then the call price $C=\mathbb{E}[(S_T-K)^+]=\mathbb{E}\big[\mathbb{E}[(S_T-K)^+|\mathcal{G}_T]\big]$.

Denote $q(y)$ the risk-neutral density of log-price $Y_T=\log(S_T)$ conditioned on $\mathcal{G}_T$. Writing the log-strike price as $k=\log(K)$, we have the conditional call price

$$c(k)=\mathbb{E}[(e^{Y_T}-K)^+|\mathcal{G}_T]=\int_k^{\infty}(e^y-e^k)q(y)dy. \tag{3.4.6}$$

Note that as $k\to-\infty$, $c(k)$ goes to a positive constant, so it is not square integrable and we can not directly apply Fourier transform on $c(k)$. To handle this, we apply Fourier transform on a modified call option instead. Using a damping coefficient $\alpha>0$, we define the modified function by $c^\alpha(k)=e^{\alpha k}c(k)$. Assume that $\mathbb{E}[(S_T)^{\alpha+1}]<\infty$. The Fourier transform of $c^\alpha(k)$ is

$$\begin{aligned}\psi^\alpha(u)&=\mathcal{F}(c^\alpha(k))=\int_{-\infty}^{+\infty}e^{iuk}c^\alpha(k)dk=\int_{-\infty}^{+\infty}e^{iuk}\int_k^{+\infty}e^{\alpha k}(e^y-e^k)q(y)dydk\\&=\int_{-\infty}^{+\infty}q(y)\int_{-\infty}^{y}e^{iuk}(e^{y+\alpha k}-e^{\alpha+\alpha k})dkdy\\&=\int_{-\infty}^{+\infty}q(y)e^{(1+\alpha+iu)y}(\frac{1}{\alpha+iu}-\frac{1}{\alpha+1+iu})dy\\&=\frac{\phi(u-(\alpha+1)i)}{\alpha^2+\alpha-u^2+i(2\alpha+1)u}.\end{aligned}$$

Now performing the inverse Fourier transform on $\psi^\alpha(u)$, we have

$$c(k)=e^{-\alpha k}c^\alpha(k)=\frac{e^{-\alpha k}}{2\pi}\int_{-\infty}^{+\infty}e^{-iuk}\psi^\alpha(u)du=\frac{e^{-\alpha k}}{\pi}\int_0^{+\infty}e^{-iuk}\psi^\alpha(u)du. \tag{3.4.7}$$

Notice that we can then use FFT to calculate (3.4.7) quickly and efficiently. To see it, we discretize u and k as $u_j=j\eta$ for $j=0,1,\cdots L-1$, and $k_n=-\frac{L\xi}{2}+\xi n$, for $n=0,1,\cdots L-1$. The sizes of the discretization satisfy $\xi\eta=\frac{2\pi}{L}$, and $L=2^d$ for some big integer d. To gain more accuracy, Simpson's rule for weightings are also incorporated in the summation. The call option price can be computed as

$$c(k_n) = \frac{e^{-\alpha k_n}}{\pi} \sum_{j=0}^{L-1} e^{-i\frac{2\pi}{N}jn} e^{i\frac{N\xi}{2}u_j} \psi^{\alpha}(u_j) \frac{\eta}{3}(3 + (-1)^{j+1} + \delta_j),$$

and $\hat{c} = c(k_{\frac{N}{2}})$ gives the price of the at-the-money call option. Denote $C^{(k)}$ to be the conditional call price for the kth path. The Monte-Carlo price can be calculated similarly as in (3.4.3).

While feasible, the FFT method seems computationally intensive, since one needs to either perform FFT for each path then average, or do one FFT after averaging the conditional transform. Besides, the bias in the final estimator introduced by the FFT error is unknown.

3.4.3 Method II: Polynomial Approximation method

Alternatively, to price a financial derivative with final payoff $f(S_T)$, we can approximate the payoff function with polynomials, then use techniques in Section 3.4.1 to get the moments of the stock, and hence estimate the price. With this approach, it is easy to characterize the magnitude of the bias of simulation estimator since the approximation error is known. Furthermore, since the evaluation is much faster, we can generate more paths to get more accurate results.

Schoutens [91], Carr and Schoutens [29] and others use linear combinations of power payoffs to approximate a payoff function $f(x)$ in a least square sense. The Weierstrass approximation theorem guarantees the existence of some polynomial which can uniformly approximate a continuous function defined on a finite interval as closely as desired. However, as seen in Carr and Schoutens [29] and mentioned in Royston and Altman [90], using conventional power polynomials with higher degree can lead to wilder oscillation, the "Runge's phenomenon". In this case, higher degree does not necessarily lead to a better fit. One way to reduce this oscillation is

to extend the approximation to fractional polynomials. Fractional polynomials are linear combinations of x^p where p can be a real number. They provide more flexible fit than conventional polynomials and can obtain similar accuracy with a smaller degree, as shown in Royston and Altman [89] and Royston and Altman [90]. However, little is known about the rate of convergence or how to choose the fractional powers optimally.

We, instead, use Chebyshev polynomials. Chebyshev series expansion can minimize the Runge's phenomenon and provide an approximation that is close to the best polynomial for a continuous function under the maximum norm (see Mason and Handscomb [76], Boyd [26], or more explicitly in the introduction of Pachón and Trefethen [83]). The Chebyshev polynomials (of the first kind) $T_n(x)$ are defined by the recurrence relation $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$, with $T_0(x) = 1, T_1(x) = x$. They are orthogonal with respect to the weight $1/\sqrt{1-x^2}$ on the interval [-1,1], and

$$< T_m(x), T_n(x) >:= \frac{1+\delta_{0n}}{\pi} \int_{-1}^{1} \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}} = \delta_{mn}.$$

A function f on [-1,1] can then be represented by Chebyshev expansion $f(x) = \sum_{n=1}^{\infty} a_n T_n(x)$, where a_n is given by the inner product $< f(x), T_n(x) >$. If a function f is $(m-1)$-times differentiable almost everywhere with $(m-1)$-st derivative of bounded variation, then it can be approximated by a M-degree Chebyshev series $\sum_{n=1}^{M} a_n T_n(x)$ with maximum absolute error of order $O(M^{-m})$ (see Theorem 2.1(ii) of Battles and Trefethen [14]).

Any finite range $0 \le y \le a$ can be transformed to the basis range $-1 \le x \le 1$ with change of variable $y = a(x+1)/2$. In our case, we approximate the call option payoff $\bar{f}(y) = (y-1)^+$ on truncated domain $[0, a]$ with $a = 4$ and transform $y = 2(x+1)$. Write $\bar{T}_n(y) = T_n(x), \bar{f}(y) = f(x)$, then $\bar{T}_n(y)$ can be made explicit and so does the Chebyshev expansion of $\bar{f}(y) \approx P_M(y) = \sum_{n=1}^{M} a_n \bar{T}_n(y)$. (Without loss of generality,

the strike price can be rescaled to unity as $\mathbb{E}[(S_T - K)^+] = K\mathbb{E}[(\bar{S}_T - 1)^+]$ where $\bar{S}$ has similar dynamic as S and $\bar{S}_0 = S_0/K$. The call price will be rescaled by a factor of K accordingly.) Since the payoff function of a call option is piecewise differentiable, the maximum absolute error is at most of order $O(1/M)$.

Assume the stock price has finite moments. The truncated domain $[0, a]$ can be chosen such that the error is negligible outside the region. Here our choice of $a = 4$ is more than three standard derivations of stock price. Of course we can pick an even bigger region $[0, b]$ with $b > a$ to obtain smaller error due to the truncation of domain, but the approximation error using the scaled Chebyshev basis $T_n(x)$ will generally get bigger. There is a trade-off in the choice of a.

Once the Chebyshev expansion is determined, we can rearrange the terms and get $\bar{f}(S) = (S-1)^+ \approx \sum_{n=0}^{M} a_n \bar{T}_n(S) = \sum_{n=0}^{M} a_n^* S^n$. The conditional call price for the kth path is then given by

$$C^{(k)} \approx \mathbb{E}[P_M(S_t)|\mathcal{G}_T] = \sum_{j=1}^{M} a_j^* M_T(-j, 0),$$

and (3.4.3) gives the Monte-Carlo estimation of the un-discounted option price.

3.5 Numerical Results

We compare the exact method with the discretization method described in Section 3.2.2. We price a European call option on the stock. We consider the following intensity specification:

$$\lambda_t^1 = c_1 V_t, \quad \lambda_t^2 = V_t g(J_t^y, J_t^v), \quad \lambda_t^3 = c_3 V_t. \tag{3.5.1}$$

where c_1 and c_3 are constants, and $g(\cdot, \cdot)$ is an arbitrary non-negative function. We assume that the initial value V_0 for the variance process has a Gamma distribution, its stationary distribution. The diffusion parameters $\kappa, \theta, \sigma, r, \rho$ are constants. The

initial log-price of stock Y_0 is independent of V_0, and could either be a fixed value or follow some distribution. The jump sizes are chosen to be independent of state processes. Assumption (A) is satisfied and strong unique solution exists for SDE (3.2.1). In this formulation, the filter $M_t(z_1, z_2)$ and projected intensities $h_{i,n}(t)$ in Algorithm 3.4.1 can be computed exactly in closed-form or approximated in closed-form as in Appendix A; see Propositions A.1 and A.2.

Under this specification, the compensator in (2.1) becomes $A_t = \lambda_t^1 \bar{\mu}_{y,o} + \lambda_t^3 \bar{\mu}_{y,c}$, where $\bar{\mu}_{y,o} = \mathbb{E}[\exp(\Delta^y) - 1]$ is the average jump size of stock price due to type I jumps (jumps in only Y_t), and $\bar{\mu}_{y,c} = \mathbb{E}[\exp(\Delta^{y,c}) - 1]$ is the average jump size of stock price due to type III jumps (common jumps). In addition, $\mu_0 = r, \mu_1 = 0, \mu_2 = \frac{1}{2} + c_1\bar{\mu}_{y,o} + c_3\bar{\mu}_{y,c}$ in equation (3.3.8) are also constants. The ODE system (3.3.8) now has closed-form solutions,

$$\begin{aligned} b_1(s,t,\mathbf{J},\mathbf{z}) &= z_1 \\ b_2(s,t,\mathbf{J},\mathbf{z}) &= \frac{z_2(C - Be^{\gamma(t-s)}) - \frac{BC}{\sigma^2}(e^{\gamma(t-s)} - 1)}{\sigma^2 z_2(e^{\gamma(t-s)} - 1) + Ce^{\gamma(t-s)} - B}, \\ a(s,t,\mathbf{J},\mathbf{z}) &= -rz_1(t-s) + \frac{2\kappa\theta}{\sigma^2}\log\left[\frac{2\gamma\exp[\frac{1}{2}(\kappa+\gamma+\rho\sigma z_1)(t-s)]}{\sigma^2 z_2(e^{\gamma(t-s)} - 1) + Ce^{\gamma(t-s)} - B}\right], \end{aligned} \tag{3.5.2}$$

with $\gamma = \sqrt{(\kappa+\rho\sigma z_1)^2 - \sigma^2 z_1(z_1+2\mu_2) + 2\sigma^2[g(j^y, j^v) + c_1 + c_3]}$, $B = k + \rho\sigma z_1 - \gamma$, and $C = k + \rho\sigma z_1 + \gamma$. The filter $M_t(z_1, z_2)$ in Algorithm 3.4.1 is hence

- for $t \in [T_n, T_{n+1})$,

$$M_t(z_1, z_2) = \frac{e^{a(T_n,t,\mathbf{J}_{T_n},\mathbf{z})} M_{T_n}(\mathbf{b}(T_n, t, \mathbf{J}_{T_n}, \mathbf{z})}{e^{a(T_n,t,\mathbf{J}_{T_n},\mathbf{0})} M_{T_n}(\mathbf{b}(T_n, t, \mathbf{J}_{T_n}, \mathbf{0})},$$

- at T_{n+1}

$$M_{T_{n+1}}(z_1, z_2) = e^{-z_1\Delta J^y_{n+1} - \delta z_2 \Delta J^v_{n+1}} \frac{\partial_{v_2}\left[e^{a(T_n,T_{n+1},\mathbf{J}_{T_n},\mathbf{v})} M_{T_n}(\mathbf{b}(T_n, T_{n+1}, \mathbf{J}_{T_n}, \mathbf{v}))\right]\Big|_{\mathbf{v}=(z_1,z_2)}}{\partial_{v_2}\left[e^{a(T_n,T_{n+1},\mathbf{J}_{T_n},\mathbf{v})} M_{T_n}(\mathbf{b}(T_n, T_{n+1}, \mathbf{J}_{T_n}, \mathbf{v}))\right]\Big|_{\mathbf{v}=(0,0)}},$$

The recursive formulas for projected intensity are

$$h_n(t) = -(c_1 + c_3 + g(\mathbf{J}_{T_n}))\frac{\partial_{z_2}\left[e^{a(T_n,t,(0,z_2),\mathbf{J}_{T_n})}M_{T_n}(0, b_2(T_n,t,(0,z_2),\mathbf{J}_{T_n}))\right]|_{z_2=0}}{e^{a(T_n,t,\mathbf{0},\mathbf{J}_{T_n})}M_{T_n}(0, b_2(T_n,t,\mathbf{0},\mathbf{J}_{T_n}))}\ , \tag{3.5.3}$$

and the probabilities $p_i = \mathbb{P}(\xi_{n+1} = i)$ for the jump-type process ξ at T_{n+1} are $p_i = c_i/[c_1 + c_3 + g(J^y_{T_n}, J^v_{T_n})]$, $i = 1, 3$, and $p_2 = 1 - p_1 - p_3$.

3.5.1 Jumps in price only.

Suppose $\lambda^2_t = \lambda^3_t = \delta = 0$ in (3.2.1). The intensity of price jumps is $\lambda^y_t = \lambda^1_t = c_1 V_t$. The jump sizes have a log-normal distribution. In this setting, the filter can be computed explicitly, see Proposition 3.7.1. We also obtain a semi-analytical value for the option price from the results of Duffie et al. [36], which will facilitate the evaluation of the simulation estimators.

The parameter values for numerical computations (see Table 3.1) are taken from Duffie et al. [36], and were found by minimizing the mean square error for market option prices on the S&P 500 on November 2, 1993. (Duffie et al. [36] assume a constant jump intensity equal to 0.11. We select the intensity parameter c_1 such that the average number of jumps per year is around 0.11.)

Methods I and II are configured as follows. We choose $\alpha = 1.5$, $L = 2^{12}$, and $\eta = 600/L$ in the calculation for FFT in Method I. The result is not sensitive to the choice of α,L, and η as long as they are appropriate. In Method II, we choose the degree of Chebyshev polynomials to be $M = 26$. The Chebyshev expansion approximates the payoff function $(S-1)^+$ with maximum absolute error of 0.02 and mean absolute error of around 0.002 on interval $[0, 4]$. The strike price is rescaled to unity as $\mathbb{E}[(S_T - K)^+] = K\mathbb{E}[(\bar{S}_T - 1)^+]$ where $\bar{S}$ has same dynamic as S and $\bar{S}_0 = S_0/K$.

To further reduce the error of approximation in method II, we could correct the error using "benchmark adjustment". Note the option price $C = \mathbb{E}[f(x)|\mathcal{G}_T] = \mathbb{E}[P^*(x)|\mathcal{G}_T] + err$, where $err = \mathbb{E}[f(x)|\mathcal{G}_T] - \mathbb{E}[P^*(x)|\mathcal{G}_T]$. We can approximate the error term with $err \approx \mathbb{E}^0[f(x)|\mathcal{G}_T] - \mathbb{E}^0[P^*(x)|\mathcal{G}_T]$ where $\mathbb{E}^0[\cdot]$ is some computable expectation under a simpler measure close to the original measure. For example, it could be the expectation when no jump occurs. That is, we perform option valuation with a single FFT and a single polynomial approximation using M_T with zero total number of jumps, then store the difference as the value err for benchmark adjustment. These single valuations are fast and generally complete in less than 0.02 second. We denote this method "adjusted Method II".

To evaluate an estimator of the option price, we consider its root mean square error (RMSE), given by $\sqrt{\text{Bias}^2 + \text{SE}^2}$. The standard error (SE) is estimated as the sample standard deviation of the simulation output divided by the square root of the number of trials. The bias is estimated using a large number of trials to estimate the expectation of the estimator, and then taking the difference with the true value.

The true value can be computed using the results of Duffie et al. [36]. They give the explicit formula for the characteristic function of terminal stock price under affine setting with fixed initial values (Y_0, V_0). Using properties of Gamma distributions, the characteristic function of stock price with Gamma distributed initial value V_0 can also be obtained. This characteristic function can be used here to compute the true option price. Below we give the details on how to get the true price.

According to (2.4)-(2.6) of Duffie et al. [36], the transform function given V_0 is given by

$$\mathbb{E}[e^{uY_T}|Y_t = y, V_t = v] = e^{\alpha(t)+\beta_1(t)y+\beta_2(t)v},$$

where $\alpha(t), \beta(t)$ solves ODEs (when $\lambda^1 = c_1 V_t, \lambda^2 = \lambda^3 = 0$ and log-normal jumps in

Y_t)

$$\begin{aligned}
&\partial_t\beta_1(t) = 0\\
&\partial_t\beta_2(t) = \left(\frac{1}{2} + c_1\bar{\mu}_{y,o}\right)\beta_1(t) + \kappa\beta_2(t) - \frac{1}{2}\left[\beta_1^2(t) + 2\rho\sigma\beta_1(t)\beta_2(t) + \sigma^2\beta_2^2(t)\right] - c_1[\mathcal{J}(\beta_1(t)) - 1]\\
&\partial_t\alpha(t) = -r\beta_1(t) - \kappa\theta\beta_2(t)
\end{aligned} \tag{3.5.4}$$

with terminal condition $\beta_1(T) = u, \beta_2(T) = 0, \alpha(T) = 0$, where $\mathcal{J}(u) = \exp(\mu_{y,o}u + \frac{1}{2}\sigma_{y,o}^2u^2)$ is the jump transform and $\bar{\mu}_{y,o} = \mathcal{J}(1) - 1$. This ODE has closed-form solution and the transform of the state process is given by

$$\mathbb{E}[e^{uY_T}|Y_t = y, V_t = v] = \exp(\bar{\alpha}(T-t,u) + uy + \bar{\beta}(T-t,u)v) \tag{3.5.5}$$

where

$$\begin{aligned}
\bar{\beta}(\tau,u) &= -\frac{a(1-e^{-\gamma\tau})}{(\gamma-b) + (\gamma+b)e^{-\gamma\tau}}\\
\bar{\alpha}(\tau,u) &= ru\tau + \frac{2\kappa\theta}{\sigma^2}\log\left(\frac{2\gamma e^{-(\gamma+b)\tau/2}}{(\gamma-b) + (\gamma+b)e^{-\gamma\tau}}\right)
\end{aligned} \tag{3.5.6}$$

where $b = \sigma\rho u - \kappa,\ a = u(1-u) + 2c_1\bar{\mu}_{y,o}u - 2c_1[\mathcal{J}(u) - 1],\ \gamma = \sqrt{b^2 + \sigma^2 a}$. Hence $\mathbb{E}[\exp(uY_T)|Y_0, V_0] = \exp(\bar{\alpha}(T,u) + uY_0 + \bar{\beta}(T,u)V_0)$. Since V_0 follows gamma distribution with scale parameter $\sigma^2/2\kappa$ and shape parameter $2\kappa\theta/\sigma^2$, we have the characteristic function of log-price as

$$\phi(u) = \mathbb{E}[\exp(iuY_T)|Y_0] = \exp(\bar{\alpha}(T,iu) + iuY_0)(1 - \bar{\beta}(T,iu)\sigma^2/2\kappa)^{-2\kappa\theta/\sigma^2}. \tag{3.5.7}$$

We then apply FFT on the characteristic function of log-price to get the true value of vanilla call option price.

The simulation experiments were performed on a desktop computer running Windows 7 with 3 GB RAM. The codes were written in Matlab R2009b (Version 7.9.0). Motivated by the results of Duffie and Glynn [35], the number of discretization time steps was set to the square root of the number of simulation trials.

Table 3.1 reports the simulation results. The bias column is estimated using 6,553,600 trails. We see that the exact simulation method leads to much lower variance, as expected, than the Euler discretization method. In addition, the bias introduced by FFT in Method I is negligible. In Method II, the bias introduced by approximation using finite number of polynomials is bounded by the approximation error, and decreases a lot after benchmark adjustment as shown in "Method II adjusted". Figure 3.1 shows the convergence of the two methods graphically. The curve for Method II flattens because the SE column with more than 102,400 trails is smaller than its bias -0.007. After benchmark adjustment, the bias in Method II was reduced and the method converges. The projection method, implemented either in Method I or adjusted Method II, achieves the optimal square-root convergence, with the error of estimator decreases at the rate $O(1/\sqrt{t})$ for computational budget t. The steeper slopes of the projection methods show faster convergence rate than that of Euler method. All projection methods also outperform the discretization scheme in terms of absolute errors: for a given computational budget, they generate smaller RMS errors. Among the projection method, adjusted Method II performs the best since it evaluates M_T at much fewer points and hence runs much faster.

3.5.2 Jumps in the price and volatility

When δ in (3.2.1) is not zero, we can still approximate the filter by an explicit expression with approximation error of order $O(\delta^2)$ as in Proposition A.2 in Appendix. Here we give an example with $\delta = 0.5$ and allow jumps in both log-price process and the variance process.

The parameter values in the numerical example (see Table 3.2) are taken from the third column of Table I from Duffie et al. [36]. We allow three types of jumps as follows,

- Type I: jumps in only the log-price process. The jump sizes follow a normal distribution with mean $\mu_{y,o}$ and standard-derivation $\sigma_{y,o}$=0.15. The value $\mu_{y,o}$ is chosen such that the mean jump sizes of the stock is $\bar{\mu}_{y,o} = -0.12$. The relation of the three parameters is given by $\mu_{y,o} = \log(\bar{\mu}_{y,o} + 1) - \sigma_{y,o}^2/2$. The intensity of this type of jumps is $\lambda_t^1 = c_1 V_t$.

- Type II: jumps in only the variance process: the jump sizes follow an exponential distribution with mean $\mu_{v,o} = 0.03$. The intensity of this type of jumps is $\lambda_t^2 = V_t \log(|1 - J_t^y + J_t^v|)$.

- Type III: common jumps in both processes: the jumps sizes in variance process $\Delta^{v,c}$ follow an exponential distribution with mean $\mu_{v,c} = 0.05$; conditional on $\Delta^{v,c}$, the jumps sizes in log-price $\Delta^{y,c}$ follow a normal distribution with mean $\mu_{y,c} + \rho_J \Delta^{v,c}$ and standard-derivation $\sigma_{y,c} = 0.0001$, where $\rho_J = -0.38$. The value $\mu_{y,c}$ is chosen such that the mean jump sizes of the stock is $\bar{\mu}_{y,c} = -0.10$. These parameters satisfy the equation $\mu_{y,c} = \log[(\bar{\mu}_{y,c} + 1)(1 - \rho_J \mu_{v,c})] - \sigma_{y,c}^2/2$. The intensity of this type of jumps is $\lambda_t^3 = c_3 V_t$.

In addition, we assume $V_0 \sim \text{Gamma}(2\kappa\theta/\sigma^2, \sigma^2/2\kappa)$. The value c_1, c_3 are chosen such that the average number of jumps per year is around 0.11.

At these parameter values, the projected intensity $h(t)$ is decreasing between jumps. Hence in Algorithm 3.4.2, we can take its initial value as an upper bound B_t^n, and don't need to find a value for C_t^n.

In this example, the point process J_t^v associated with the variance process V is self-exciting since its jump intensity is a function of the state process and the point process itself. The intensity $\lambda_t^2 = V_t \log(|1 - J_t^y + J_t^v|)$ suggests that more jumps occur for a higher level of volatility point process or a lower level of stock price point

process. Since the intensity is a non-linear function of point processes, it goes beyond the affine model of Duffie et al. [36].

Table 3.2 and Figure 3.2 give the simulation results and convergence diagram of RMS errors. Here both "Method I" and "Method I*" use FFT techniques as in Section 3.4.2 to compute the option price based on the filter, while "Method I" calculates the filter numerically using Algorithm 4.1 and "Method I*" approximates the filter by Proposition A.2. Similarly, both "adjusted Method II" and "adjusted Method II*" apply the benchmark adjustment after the polynomial approximation method in Section 3.4.3 to get option price based on the filter, while the filter is computed, respectively, by numerical recursion in "adjusted Method II" and by Proposition A.2 in "adjusted Method II*".

The bias column is estimated using 6,553,600 trails except for Method I/II, which use 409,600 trails due to the cost of numerical recursion to compute M_T. For a given computation budget, all projection method outperform the discretization method, except Method I which uses numerical recursive calculation together with FFT. During the numerical recursion, large proportion of the computational time was spent on solving ODE (3.3.8) for different z_1 necessary in FFT. If ODE (3.3.8) had an explicit solution, or can be solved quickly for a vector of z_1, Method I would beat discretization method. The adjusted Method II* flattens because the bias is bigger than the SE column with more than 409,600 trails. The bias comes from two places: the approximation of M_T using Proposition A.2 with error $O(\delta^2)$ and the Chebyshev polynomial approximation to the payoff function with error $O(1/M)$ where M is the degree of polynomial. If one can reduce these errors, the curve will be straight. We see that projection method still achieves the optimal square-root convergence rate even for the approximated filter using Proposition A.2. The discretization method,

on the other hand, has a slower convergence rate.

3.5.3 Extensions

Our method can be generalized further to Heston model with time-dependent diffusion coefficients. For example, piece-wise constant coefficients are widely used, as in Mikhailov and Nögel [80], Elices [39], Benhamou et al. [23], and many others, due to its flexibility in fitting market data. Our exact simulation method can easily handle models with time-dependent diffusion parameters $r(t)$, $\kappa(t)$, $\theta(t)$, and $\sigma(t)$. After all, no additional work needs to be done except for solving the equation system (3.3.8), while numerous efficient and accurate numerical ODE solvers are available. No restriction needs to be imposed on the joint distribution of initial state process (Y_0, V_0). They can be either fixed values or follow a certain joint distribution. The global jump intensities can take a general form $\lambda_t = g_0(J_t^y, J_t^v) + g_1(J_t^y, J_t^v)V_t + g_2(J_t^y, J_t^v)Y_t$. As seen in previous example, for call option price valuation, the projection method generates smaller standard errors and smaller bias for a given number of trails than discretization method. But it runs slowly due to the intensive ODE solving and intensive evaluation at the FFT points. We suspect that the projection method in a very general case may be computational expensive. However, if we want to evaluate the price of such contracts as in section 4.1 which directly use the moments of state processes, the projection method would be superior to discretization method.

3.6 Conclusion

This chapter extends the Heston model to include state-dependent jumps and develops methods for exact simulation of the model. The exact sampling, based on a change of the filtration that describes the information flow in the point process and facilitated by filtering arguments, extends a projection method developed by

Giesecke et al. [49]. Numerical experiments demonstrate the effectiveness of the exact methods and their advantages over Euler discretization schemes. Our unbiased scheme requires a smaller computational budget to achiever a given error and has a faster convergence rate.

3.7 Appendix: Derivation of an explicit formula for the filter

Here we provide and prove an explicit formula for the filter $M_{T_{n+1}}(z_1, z_2)$ used in section 3.5.1 when the intensities take the form (3.5.1).

Proposition 3.7.1. *Denote $\phi_0(z_1) = \mathbb{E}[e^{-z_1 Y_0}]$ to be the moment generating function of Y_0. Assume that V_0 follows Gamma distribution with parameters $(2\kappa\theta/\sigma^2, \sigma^2/2\kappa)$ and the sensitivity parameter $\delta = 0$, then*

$$\begin{aligned} M_{T_n}(z_1, z_2) = &\left(\frac{F_n(z_1)}{F_n(0)}\right)^{\frac{2\kappa\theta}{\sigma^2}} e^{-z_1 r T_n} \phi_0(z_1) e^{-z_1 \sum_{k=1}^n \Delta J_n^y} \\ &\cdot (z_2 H_n(z_1) + K_n(z_1))^{-\frac{2\kappa\theta}{\sigma^2} - n} (K_n(0))^{\frac{2\kappa\theta}{\sigma^2} + n} \frac{P_n^{(z_1)}(z_2)}{P_n^0(0)} \end{aligned} \tag{3.7.1}$$

where $P_n^{(z_1)}(z_2)$ is a polynomial of z_2 with degree $n-1$, which satisfies

$$P_0(z_2) = 1, \quad P_1(z_2) = H_1$$

$$\begin{aligned} P_n(z_2) = &(U_n z_2 + W_n)^{n-2} P_{n-1}\left(\frac{R_n z_2 + S_n}{U_n z_2 + W_n}\right)\left[-\frac{2\kappa\theta}{\sigma^2} U_n (H_n z_2 + K_n) - \left(\frac{2\kappa\theta}{\sigma^2} + n - 1\right) G_n\right] \\ &+ (U_n z_2 + W_n)^{n-3} P'_{n-1}\left(\frac{R_n z_2 + S_n}{U_n z_2 + W_n}\right)(H_n z_2 + K_n) G_n, \quad n \geq 2, \end{aligned}$$

with $F_n = F_n(z_1) = D_1 D_2 \cdots D_n$, $D_n = (2\gamma)\exp(\frac{1}{2}(\kappa + \gamma + \rho\sigma z_1)(T_n - T_{n-1}))$, $G_n = R_n W_n - S_n U_n = (2\gamma)^2 e^{\gamma(T_n - T_{n-1})}$, and

$$H_n = R_n H_{n-1} + U_n K_{n-1}, \;\; H_0 = \sigma^2/2\kappa,$$

$$K_n = S_n H_{n-1} + W_n K_{n-1}, \;\; K_0 = 1,$$

where

$$R_n = R_n(z_1) = C_n - B_n e^{\gamma(T_n - T_{n-1})},$$

$$S_n = S_n(z_1) = (2g(J_{1,n}, J_{2,n}) + 2c_1 + 2c_2 - z_1(z_1 + 2\mu_2))(e^{\gamma(T_n - T_{n-1}} - 1),$$

$$U_n = U_n(z_1) = \sigma^2(e^{\gamma(T_n - T_{n-1})} - 1),$$

$$W_n = W_n(z_1) = C_n e^{\gamma(T_n - T_{n-1})} - B_n,$$

with $\gamma = \gamma_n = \sqrt{(\kappa + \rho\sigma z_1)^2 - \sigma^2 z_1(z_1 + 2\mu_2) + 2\sigma^2[g(J^y_{T_{n-1}}, J^v_{T_{n-1}}) + c_1 + c_2]}$, $\quad B_n = k + \rho\sigma z_1 - \gamma, \quad C_n = k + \rho\sigma z_1 + \gamma.$

Proof. First of all, denoting $D_n = 2\gamma \exp[\frac{1}{2}(\kappa + \gamma + \rho\sigma z_1)(T_n - T_{n-1})]$, we can re-write the functions $a(\cdot)$ and $b_2(\cdot)$ in (3.5.2) as [2]

$$a(T_{n-1}, T_n, (z_1, z_2)) = -z_1 r(T_n - T_{n-1}) + \frac{2\kappa\theta}{\sigma^2} \log\left(\frac{D_n}{U_n z_2 + W_n}\right) ,$$

$$b_2(T_{n-1}, T_n, (z_1, z_2)) = \frac{z_2 R_n + S_n}{z_2 U_n + W_n} ,$$

and

$$\partial_{z_2} a(T_{n-1}, T_n, (z_1, z_2)) = -\frac{2\kappa\theta}{\sigma^2} \frac{U_n}{U_n z_2 + W_n} ,$$

$$\partial_{z_2} b_1(T_{n-1}, T_n, (z_1, z_2)) = 0 ,$$

$$\partial_{z_2} b_2(T_{n-1}, T_n, (z_1, z_2)) = \frac{R_n W_n - S_n U_n}{(U_n z_2 + W_n)^2} = \frac{G_n}{(U_n z_2 + W_n)^2} .$$

For $n = 0$, $M_0(z_1, z_2) = \phi_0(z_1)(1 + z_2)^{-\frac{2\kappa\theta}{\sigma^2}}$. For $n > 0$,

$$M_{T_n}(z_1, z_2) = e^{-z_1 \Delta J^y_n} \frac{\partial_{z_2} e^{a(z_1, b_2(T_{n-1}, T_n, (z_1, z_2)))} M_{T_{n-1}}(z_1, b_2(T_{n-1}, T_n, (z_1, z_2)))}{\partial_{z_2} e^{a(z_1, b_2(T_{n-1}, T_n, (z_1, z_2)))} M_{T_{n-1}}(z_1, b_2(T_{n-1}, T_n, (z_1, z_2)))|_{(z_1, z_2) = (0,0)}}. \tag{3.7.2}$$

[2] to get shorter notations, we make the dependency of $(J^y_{T_{n-1}}, J^v_{T_{n-1}})$ implicit here.

Assume claim (3.7.1) holds for $n-1$, then the numerator of expression in (3.7.2) is

$$
\begin{aligned}
&\partial_{z_2} e^{a(T_{n-1},T_n,(z_1,z_2))} M_{T_{n-1}}(z_1, b_2(T_{n-1},T_n,(z_1,z_2))) \\
=&e^{a(T_{n-1},T_n,(z_1,z_2))} M_{T_{n-1}}(z_1, b_2(T_{n-1},T_n,(z_1,z_2))) \\
&\cdot \left[\partial_{z_2} a(T_{n-1},T_n,(z_1,z_2)) + \partial_v \log(M_{T_{n-1}}(z_1,v))|_{v=b_2(T_{n-1},T_n,(z_1,z_2))} \partial_{z_2} b_2(T_{n-1},T_n,(z_1,z_2))\right] \\
=&C_0 e^{-rT_n} e^{-z_1 Y_0 - z_1 \sum_{k=1}^{n-1} \Delta J_k^y} \left(\frac{D_n}{z_2U_n+W_n}\right)^{\frac{2\kappa\theta}{\sigma^2}} (F_{n-1})^{\frac{2\kappa\theta}{\sigma^2}} \left[\frac{z_2R_n+S_n}{z_2U_n+W_n} H_{n-1} + K_{n-1}\right]^{-\frac{2\kappa\theta}{\sigma^2}-n+1} \\
&\cdot P_{n-1}\left(\frac{z_2R_n+S_n}{z_2U_n+W_n}\right)\Big[-\frac{2\kappa\theta}{\sigma^2}\frac{U_n}{z_2U_n+W_n} \\
&\qquad + \left(\frac{-(\frac{2\kappa\theta}{\sigma^2}-n+1)H_{n-1}}{\frac{z_2R_n+S_n}{z_2U_n+W_n}H_{n-1}+K_{n-1}} + \frac{P'_{n-1}(\frac{z_2R_n+S_n}{z_2U_n+W_n})}{P_{n-1}(\frac{z_2R_n+S_n}{z_2U_n+W_n})}\right)\frac{G_n}{(z_2U_n+W_n)^2}\Big] \\
=&C_0 e^{-rT_n} e^{-z_1 Y_0} e^{-z_1\sum_{k=1}^{n-1}\Delta J_k^y} (F_n)^{\frac{2\kappa\theta}{\sigma^2}} (z_2H_n+K_n)^{-\frac{2\kappa\theta}{\sigma^2}-n+1} (z_2U_n+W_n)^{n-2} \\
&\cdot P_{n-1}\left(\frac{z_2R_n+S_n}{z_2U_n+W_n}\right)\left[-\frac{2\kappa\theta}{\sigma^2}U_n + \left(\frac{-(\frac{2\kappa\theta}{\sigma^2}-n+1)H_{n-1}}{z_2H_n+K_n} + \frac{P'_{n-1}(\frac{z_2R_n+S_n}{z_2U_n+W_n})}{P_{n-1}(\frac{z_2R_n+S_n}{z_2U_n+W_n})(z_2U_n+W_n)}\right)G_n\right] \\
=&C_0 e^{-rT_n} e^{-z_1 Y_0} e^{-z_1\sum_{k=1}^{n-1}\Delta J_k^y} (F_n)^{\frac{2\kappa\theta}{\sigma^2}} (z_2H_n+K_n)^{-\frac{2\kappa\theta}{\sigma^2}-n} P_n(z_2).
\end{aligned}
$$

Here, C_0 is some constant, and

$$
\begin{aligned}
P_n(z_2) &= (z_2U_n+W_n)^{n-2}(z_2H_n+K_n)\Big[-\frac{2\kappa\theta}{\sigma^2}U_n \\
&\qquad + \left(\frac{-(\frac{2\kappa\theta}{\sigma^2}-n+1)}{z_2H_n+K_n} + \frac{P'_{n-1}(\frac{z_2R_n+S_n}{z_2U_n+W_n})}{P_{n-1}(\frac{z_2R_n+S_n}{z_2U_n+W_n})(z_2U_n+W_n)}\right)G_n\Big] \\
&= (U_nz_2+W_n)^{n-2}P_{n-1}(\frac{R_nz_2+S_n}{U_nz_2+W_n})\Big[-\frac{2\kappa\theta}{\sigma^2}U_n(H_nz_2+K_n) - (\frac{2\kappa\theta}{\sigma^2}+n-1)G_n\Big] \\
&\quad + (U_nz_2+W_n)^{n-3}P'_{n-1}(\frac{R_nz_2+S_n}{U_nz_2+W_n})(H_nz_2+K_n)G_n.
\end{aligned}
$$

Note that $P_{n-1}(z_2)$ is a polynomial of z_2 with degree $n-2$, so $(U_nz_2+W_n)^{n-2}P_{n-1}(\frac{R_nz_2+S_n}{U_nz_2+W_n})$ is a polynomial of z_2 with degree $n-2$, and $(U_nz_2+W_n)^{n-3}P'_{n-1}(\frac{R_nz_2+S_n}{U_nz_2+W_n})$ is a polynomial with degree $n-3$. In a word, $P_n(z_2)$ is a polynomial of z_2 with degree $n-1$. Hence $M_{T_n}(z_1,z_2)$ have the expression (3.7.1).

Proposition 3.7.2 (Approximated Exact Simulation when $\delta \neq 0$ in (3.2.1))**.** *Given $\delta << 1$, and define $\widetilde{M}_{T_n}(z_1,z_2)$ to be $(1-\delta\Delta J_n^v z_2)$ multiplied by the one given in the*

formula (3.7.1), then the error $|M_{T_n}(z_1, z_2) - \widetilde{M}_{T_n}(z_1, z_2)|$ *is of order* $O(\delta^2)$.

Proof. The proof is similar to the proof of Proposition A.2 in Giesecke et al. [49].

Method	Trials	Time-steps	Estimator	Bias	SE	RMSE	Time(sec)
Method I	1,600	N/A	7.04246	2.48E-4	0.05185	0.05185	2.15
	6,400	N/A	7.04324	2.48E-4	0.02662	0.02662	8.47
	25,600	N/A	7.04352	2.48E-4	0.01353	0.01353	34.64
	102,400	N/A	7.02434	2.48E-4	0.00650	0.00650	137.24
	409,600	N/A	7.02312	2.48E-4	0.00331	0.00331	559.12
	1,638,400	N/A	7.02069	2.48E-4	0.00163	0.00165	2149.50
	6,553,600	N/A	7.02022	2.48E-4	0.00082	0.00086	8571.25
Method II	1,600	N/A	7.05777	-0.00706	0.04929	0.04979	0.29
	6,400	N/A	7.00064	-0.00706	0.02800	0.02887	1.15
	25,600	N/A	7.01282	-0.00706	0.01306	0.01485	4.59
	102,400	N/A	7.01575	-0.00706	0.00660	0.00966	18.34
	409,600	N/A	7.01086	-0.00706	0.00329	0.00779	75.61
	1,638,400	N/A	7.01629	-0.00706	0.00163	0.00724	296.53
	6,553,600	N/A	7.01292	-0.00706	0.00082	0.00710	1191.19
Adjusted Method II	1,600	N/A	7.06473	-8.85E-05	0.04929	0.04929	0.31
	6,400	N/A	7.00761	-8.85E-05	0.02800	0.02800	1.17
	25,600	N/A	7.01978	-8.85E-05	0.01306	0.01307	4.60
	102,400	N/A	7.02272	-8.85E-05	0.00660	0.00660	18.36
	409,600	N/A	7.01783	-8.85E-05	0.00329	0.00329	75.62
	1,638,400	N/A	7.02326	-8.85E-05	0.00163	0.00164	296.55
	6,553,600	N/A	7.01989	-8.85E-05	0.00082	0.00082	1191.20
Discretization	1,600	40	7.05283	-0.00738	0.19070	0.19074	0.46
	6,400	80	7.09184	-0.01013	0.09508	0.09510	3.01
	25,600	160	6.97298	-0.00242	0.04690	0.04802	21.85
	102,400	320	7.00894	-0.00262	0.02369	0.02407	166.25
	409,600	640	7.00193	0.00153	0.01186	0.01187	1291.06
	1,638,400	1280	7.01313	-0.00081	0.00603	0.00609	9432.17
	6,553,600	2560	7.01846	-0.00151	0.00297	0.00333	80421.19

Table 3.1: Simulation result for a European call option (un-discounted price). Parameters: $S_0 = 100$, $K = 100$, $r = 3.19\%$, $\kappa = 3.99$, $\theta = 0.014$, $\sigma = 0.27$, $\rho = -0.79$, $\bar{\mu}_{y,o} = -0.12$, $\sigma_{y,o} = 0.15$, $c_1 = 7$, $T = 1$ year, and $V_0 \sim \text{Gamma}(2\kappa\theta/\sigma^2, \sigma^2/2\kappa)$. True option price (un-discounted) is 7.01998. (Method I: projection method A.1 with FFT; Method II: projection method A.1 with Chebyshev polynomial approximation to payoff. Adjusted Method II: Method II with benchmark adjustment.)

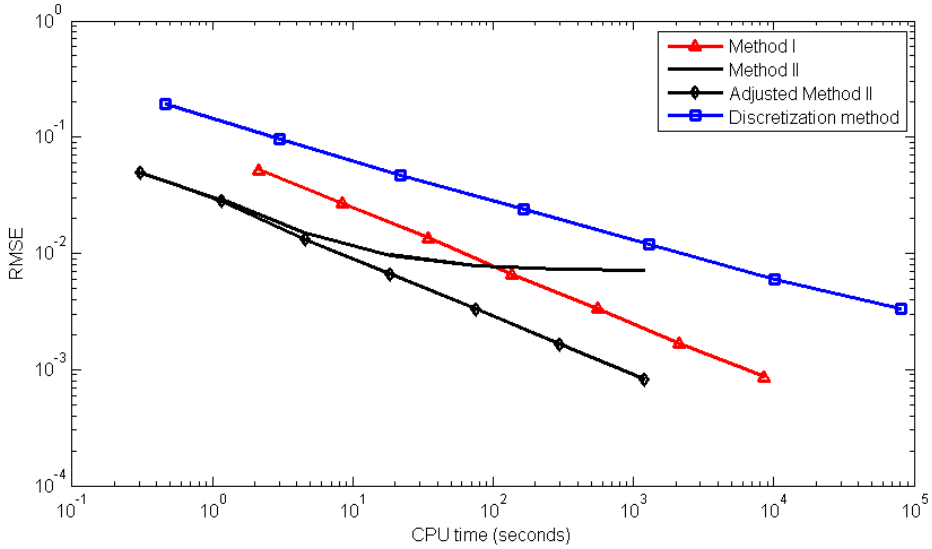


Figure 3.1: Convergence of the exact method and Euler method. Parameter values as in Table 3.1. Exact method performs much better than Euler method. The curve for Method II flattens because the SE column with more than 102,400 trails is smaller than its bias -0.007. After benchmark adjustment, the Adjusted Method II converges.

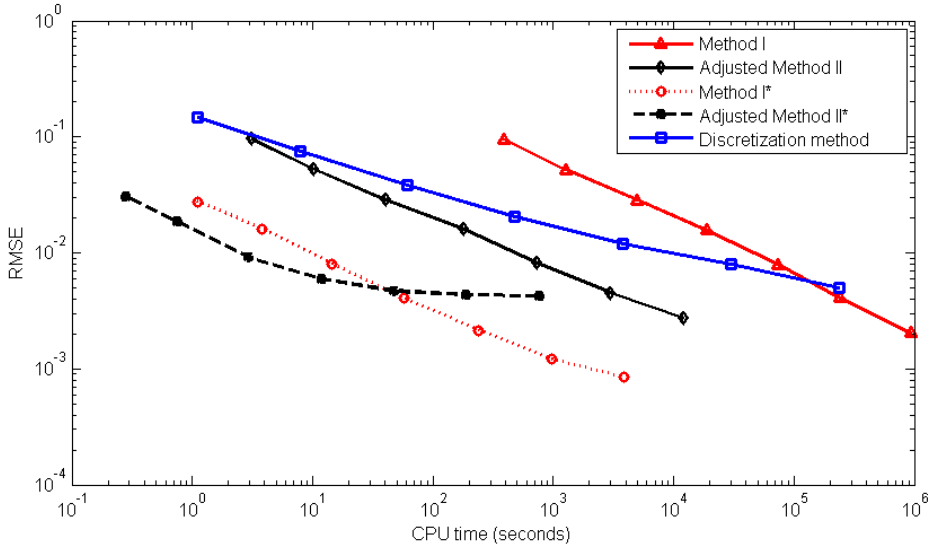


Figure 3.2: Convergence of the projection method and Euler method in terms of Root Mean Square Error(RMSE). Parameter values as in Table 3.2. The adjusted Method II* flattens because the bias is bigger than the SE column with over 409,600 trails.

Method	Trails	Time-steps	Estimation	Bias	SE	RMSE	Time(sec)
Method I	100	N/A	5.76222	0	0.09432	0.09432	390.32
	400	N/A	5.77321	0	0.05151	0.05151	1274.32
	1600	N/A	5.80043	0	0.02820	0.02820	5002.32
	6400	N/A	5.76599	0	0.01562	0.01562	18989.40
	25,600	N/A	5.77744	0	0.00785	0.00785	73858.13
	102,400	N/A	5.78101	0	0.00404	0.00404	242716.94
	409,600	N/A	5.77432	0	0.00204	0.00204	944714.40
Adjusted Method II	100	N/A	5.76154	0.00185	0.09512	0.09514	3.10
	400	N/A	5.76842	0.00185	0.05257	0.05260	10.19
	1,600	N/A	5.79650	0.00185	0.02847	0.02853	40.61
	6,400	N/A	5.76132	0.00185	0.01573	0.01583	181.91
	25,600	N/A	5.77310	0.00185	0.00793	0.00814	737.01
	102,400	N/A	5.77657	0.00185	0.00408	0.00448	2994.34
	409,600	N/A	5.77617	0.00185	0.00202	0.00274	12010.97
Method I*	1,600	N/A	5.76188	-0.00068	0.02751	0.02752	1.10
	6,400	N/A	5.73967	-0.00068	0.01611	0.01612	3.83
	25,600	N/A	5.77708	-0.00068	0.00799	0.00802	14.38
	102,400	N/A	5.77445	-0.00068	0.00401	0.00407	57.89
	409,600	N/A	5.77365	-0.00068	0.00203	0.00214	241.19
	1,638,400	N/A	5.77217	-0.00068	0.00101	0.00122	972.26
	6,553,600	N/A	5.77364	-0.00068	0.00051	0.00085	3871.69
Adjusted Method II*	1,600	N/A	5.75818	-0.00422	0.03021	0.03050	0.28
	6,400	N/A	5.76784	-0.00422	0.01824	0.01872	0.75
	25,600	N/A	5.77913	-0.00422	0.00818	0.00921	2.92
	102,400	N/A	5.76183	-0.00422	0.00413	0.00591	11.78
	409,600	N/A	5.77155	-0.00422	0.00204	0.00469	47.02
	1,638,400	N/A	5.77091	-0.00422	0.00102	0.00434	187.99
	6,553,600	N/A	5.77010	-0.00422	0.00051	0.00425	759.97
Discretization	1,600	40	5.61963	0.00826	0.14747	0.14770	1.11
	6,400	80	5.77398	0.00801	0.07534	0.07576	7.91
	25,600	160	5.73435	0.00808	0.03733	0.03819	61.30
	102,400	320	5.78812	0.00800	0.01892	0.02054	479.19
	409,600	640	5.78257	0.00712	0.00945	0.01184	3796.70
	1,638,400	1280	5.78412	0.00641	0.00473	0.00797	29882.27
	6,553,600	2560	5.77874	0.00442	0.00236	0.00501	238420.54

Table 3.2: Simulation result for a European call option (un-discounted price). Jump occurs in both log-price process and variance process. Non-affine intensities $\lambda_t^1 = 5V_t, \lambda_t^2 = V_t \log(|1 - J_t^y + J_t^v|), \lambda_t^3 = V_t$. Parameters: $\bar{\mu}_{y,o} = -0.12, \mu_{v,o} = 0.03, \mu_{v,c} = 0.05, \rho_J = -0.38, \bar{\mu}_{y,c} = -0.1$, and others are same as in Table 1. True price is given by 5.77432, estimated by Method I with 409,600 trails instead of 6,553,600 trails due to expensive computational cost. (Method I: recursive projection method with FFT; Method II: recursive projection method with Chebyshev polynomial approximation to payoff; Method I*/II*: use Proposition A.2 to approximate M_T.)

CHAPTER IV

Pricing Sovereign Credit Default Swaps with a Regime Switching Model

4.1 Introduction

Sovereign credit risk in the global financial market has gained much attention in recent years. Among all credit derivative instruments in emerging markets, sovereign credit default swap (CDS) contracts are considered to be the most liquid and actively traded. They serve as insurance against credit loss and oblige the seller to compensate the buyer in the event of loan default. In addition, they have full term-structures available to infer default risk from market data. Due to these reasons, sovereign CDS spreads have become popular in investigating emerging credit markets.

There has been extensive research on the sovereign credit market. For example, Edwards(1984, 1986, 2002), Collin-Dufresne et al. [32], Dooley [34], Berg and Sachs [24], Boehmer and Megginson [25], Duffie et al. [37], Zhang [98] and many others have investigated the determinants for sovereign credit spreads. Some recent studies suggest that sovereign credit spreads are linked to common global and market factors. Such research includes Kamin and Von Kleist [65], Eichengreen and Mody [38], Mauro et al. [77], Geyer et al. [48], Gonzalez-Rozada and Yeyati [53], Remolona et al. [88],Pan and Singleton [85], Ang and Longstaff [8], and Longstaff et al. [74]. In particular, Pan and Singleton [85] used a single-factor model with the default

rate following a log-normal process, estimated loss rate, and linked risk premiums to economic variables. Longstaff et al. [74] applied the framework in Pan and Singleton [85] to a larger sample of sovereigns and explored the linked risk premiums to broader sets of macroeconomic variables. Ang and Longstaff [8] paralleled the U.S state credit risk with the EMU sovereign credit risk and estimated the systemic factors as well as sovereign-specific factors of sovereign credit spreads from CDS term structures. The parameters in their model were estimated by least square method under risk-neutral measure.

The model we developed in this paper has several distinguishing features. Firstly, it admits both a systemic factor for the global market and a heterogeneous factor for each country. Our model can thus capture both systematic and country-specific risks. Secondly, it takes into account both risk-neutral and physical measures and provides a way to analyze the market price of risk. Thirdly, the model incorporates regime-switching of global economic states and captures the time-changing behavior of sovereign credit risk.

Since the pioneering work of Hamilton [58], regime-switching models have been widely used in modeling equity returns, interest rates, exchange rates and asset allocation. For examples, see Hamilton and Susmel [60], Hamilton and Lin [59],Ang and Chen (2002),Perez-Quiros and Timmermann (2000), Gu (2005), and Guidolin and Timmermann (2008b) for applications to equity returns; see Gray [54], Bekaert et al. (1997), and Ang and Bekaert [5], Li and Xu [73] for applications to the short rate and bond yields; see Ang and Bekaert [6], Ang and Bekaert [7], Guidolin and Timmermann [55, 56] for applications to asset allocations. As discussed in Hamilton [58], regime switching models effectively capture the abrupt changes or dramatic breaks in economic variables and financial data. This is the case for CDS spreads,

which behaved quite differently before and after the sub-prime crisis. The regime shifting model can capture this time-changing behavior.

In this study, we analyzed weekly CDS data with full term-structures. The data consists of one year up to ten year CDS spreads in the United States as well as in five other sovereign countries. Using maximum likelihood estimation for the regime-switching model as in Dai et al. [33], we are able to infer the model parameters from market data.

Several findings are presented in the preliminary results. We have found that the global-systemic risk factor behaves differently in high-volatility regimes compared to low-volatility regimes, in which its reverting speed and volatility are much higher in the high-volatility regimes. Furthermore, the sensitivity to systemic risk and the time series dynamics of sovereign-specific risk factors is shown to be heterogeneous across these different sovereign countries.

The remainder of this chapter is organized as follows. In section 4.2, we develop a regime-switching model for sovereign credit default swap, which features a mean-reverting global factor and mean-reverting country-specific factors. Section 4.3 describes the maximum likelihood method. Section 4.4 examines the data, presents empirical findings, and draws together our conclusions.

4.2 A Regime Switching Model

In this section, we introduce a regime shifting model for credit default swaps. The model features a decomposition of a global systemic risk factor and sovereign-specific risk factors.

4.2.1 Regime-dependent hazard rate of default

We assume that the regime variable $R(t) \in \{1, \cdots, K\}$ follows a continuous-time Markov Chain with a constant transition matrix

$$Q = \{q_{ij}\}_{\{i,j=1,\cdot,K\}}, \quad \sum_{k=1}^{K} q_{ik} = 0, \ q_{i,j\neq i} \geq 0,$$

where K is the total number of regimes. The hazard rate for each of the sovereign countries is a mixture of regimes and continuous processes given by

$$h^{R(t-)}(t) = \delta_0^{R(t-)} + \delta^{R(t-)} g(t) + z(t), \tag{4.2.1}$$

with constant parameter δ_0^k and sensitivity index δ^k for each regime k.

The regime-dependent continuous process $g(t)$ is global, representing a systemic risk factor that affect all countries; while the process $z(t)$ is sovereign-specific, representing a risk factor that is unique for each country. The factors $g(t)$ and $z(t)$ are independent, satisfying

$$dg(t) = \kappa_g^k[\theta_g^k - g(t)]dt + \sigma_g^k \sqrt{g(t)} dW_g(t), \ R(t-) = k, \tag{4.2.2}$$

$$dz(t) = \kappa_z[\theta_z - z(t)]dt + \sigma_z \sqrt{z(t)} dW_z(t), \tag{4.2.3}$$

where κ_z, θ_z, σ_z, κ_g's, θ_g's and σ_g's are constants, and W_g and W_z are independent Brownian motions under the risk-neutral measure.

4.2.2 Pricing general default-sensitive securities

Denote $G(t)$ be the diagonal matrix with k-th diagonal element $\delta_0^k + \delta^k g(t)$ and let $H(t) = G(t) + z(t)I$. Let $P(t,T)$ be the price vector associated with a payoff $P(T)$ at maturity T if no default occurs and a payoff $P^D(s)$ if it defaults at $s \in [t,T)$. The i-th element of the pricing vector $P(t,T)$ is the price of the securities when the

current regime is $R(t) = i$. It can be shown, similar to Proposition 2 in Farnsworth and Li [42], that

$$P(t,T) = \mathbb{E}\left[\int_t^T \{[Q - r(s)I - H(s)]P(s,T) + H(s)P^D(s)\}ds + P(T)\right] \tag{4.2.4}$$

Suppose that the default-free interest rate $r(t)$ is independent of regimes and z's. Denote by $P_0(t,T)$ the price of a default-free bond with maturity T. If the payoff at default does not depend on the price just before defaults, for example, if the recovery rate is based on principal, then the pricing equation can be rewritten as

$$\begin{aligned} P(t,T) = {} & P_0(t,T)\mathbb{E}_t\left[\exp\left(-\int_t^T z(a)da\right)\Phi(t,T)P(T)\right] \\ & + \int_0^T P_0(t,s)\mathbb{E}_t\left[\exp\left(-\int_t^s z(a)da\right)\Phi(t,s)[G(s) + z(s)I]P^D(s)\right]ds, \end{aligned} \tag{4.2.5}$$

where $\Phi(t,s)$ is defined as the solution to the following ordinary differential equation

$$\frac{d\Phi(t,s)}{dt} = -[Q - G(t)]\Phi(t,s), \quad t \le s \tag{4.2.6}$$

with boundary condition $\Phi(s,s) = I$.

This pricing equation is intuitive. The matrix $\Phi(t,s)$ represents the probability that the security has not default yet up to time s, the matrix $[G(s) + z(s)I]ds = H(s)ds$ represents the default probability over time period $(s, s+ds]$. Thus summing (integrating) all of the expected discount cash flow under the risk-neutral probability yields the price of the security.

Since the continuous processes $g(t)$ and $z(t)$ are assumed to be independent, we can separate the expectations with respect to $g(t)$ and $z(t)$. To obtain a closed-form expression of the pricing equation, we define

$$\Lambda_z(\tau, z(t), x, y) = \mathbb{E}_t[(x + yz(s))\exp(-\int_t^s z(a)da)]$$

where x and y are constants and $\tau = s - t$. Similarly, we define

$$\Lambda_g(\tau, g(t), \mathbf{x}, \mathbf{y}) = \mathbb{E}_t[\Phi(t,s)(\mathbf{x} + \mathbf{y}G(s))],$$

where $\mathbf{x}$ and $\mathbf{y}$ are K-dimensional constant vectors. In the next few subsections, we derive an explicit expression for these two quantities and hence for the pricing equation.

4.2.3 An explicit expression for Λ_z

Notice that the country-specific factor $z(t)$ follows the CIR process. Similar to affine term-structure arguments, we get

$$\begin{aligned}\Lambda_z(\tau, z(t), x, y) &= \mathbb{E}_t\left[[x + yz(s)]\exp\left(-\int_t^s z(a)da\right)\right] \\ &= [A_z(\tau, x, y) + C_z(\tau, x, y)z(t)]e^{B_z(\tau)z(t)},\end{aligned} \tag{4.2.7}$$

where functions A_z, B_z, and C_z satisfy a set of ordinary differential equations:

$$\begin{aligned}0 &= -\frac{dA_z}{d\tau} + \kappa_z\theta_z(C_z + A_zB_z), \\ 0 &= -\frac{dC_z}{d\tau} - (\kappa_z - \kappa_z\theta_zB_z - \sigma_z^2B_z)C_z, \\ 0 &= -\frac{dB_z}{d\tau} - \kappa_zB_z + \frac{1}{2}\sigma_z^2B_z^2 - 1,\end{aligned}$$

with initial conditions $A_z(0, x, y) = x, B_z(0, x, y) = 0, C_z(0, x, y) = y$.

This ODE can be solved in closed-form and the explicit expressions for A_z, B_z, C_z are

$$\begin{aligned}B(\tau, x, y) &= -\frac{2(e^{\gamma\tau} - 1)}{(\gamma - \kappa_z) + (\gamma + \kappa_z)e^{\gamma\tau}} \\ C(\tau, x, y) &= ye^{-\kappa_z\tau}\left(\frac{2\gamma e^{(\gamma+\kappa_z)\tau/2}}{(\gamma - \kappa_z) + (\gamma + \kappa_z)e^{\gamma\tau}}\right)^{2+\frac{2\kappa_z\theta_z}{\sigma_z^2}} \\ A(\tau, x, y) &= \left(\frac{2\gamma e^{(\gamma+\kappa_z)\tau/2}}{(\gamma - \kappa_z) + (\gamma + \kappa_z)e^{\gamma\tau}}\right)^{\frac{2\kappa_z\theta_z}{\sigma_z^2}}[x - y\kappa_z\theta_zB(\tau, x, y)]\end{aligned}$$

with $\gamma = \sqrt{\kappa_z^2 + 2\sigma_z^2}$.

Notice that $B(\tau, x, y) = B(\tau)$ does not depend on the initial values of x and y, and $C(\tau, x, y) = C(\tau, y)$ does not depend on x.

4.2.4 A closed-form approximation to Λ_g

Under the setting of the model, $\Lambda_g(\tau, \mathbf{x}, \mathbf{y})$ has no closed-form solutions. However, it has an approximation that is characterized by a set of ordinary differential equations.

By Ito's formula and dynamic programming principle, we have

$$\frac{\partial \Lambda_g(\tau, g, \mathbf{x}, \mathbf{y})}{\partial t} + \mathcal{A}\Lambda_g(\tau, g, \mathbf{x}, \mathbf{y}) + [Q - G(t)]\Lambda_g(\tau, g, \mathbf{x}, \mathbf{y}) = 0, \tag{4.2.8}$$

where $\mathcal{A}$ is the infinitesimal generator given by

$$\mathcal{A}\Lambda_g^k = \kappa_g^k(\theta_g^k - g)\frac{\partial \Lambda_g^k}{\partial g} + \frac{1}{2}g(\sigma_g^k)^2\frac{\partial^2 \Lambda_g^k}{\partial g^2}, \quad k = 1, 2, \cdots K.$$

As shown in Appendix 4.5.1, the differential equation (4.2.4) has an approximate solution of

$$\Lambda_g^k(\tau, g, \mathbf{x}, \mathbf{y}) = [A_g^k(\tau, \mathbf{x}, \mathbf{y}) + C_g^k(\tau, \mathbf{x}, \mathbf{y})g)]e^{B_g^k(\tau, \mathbf{x}, \mathbf{y})g}, \tag{4.2.9}$$

where $\tau = s - t$, and A_g^k, B_g^k, C_g^k satisfy the ODE

$$\begin{aligned}
0 &= -\frac{dA_g^k}{d\tau} + \kappa_g^k\theta_g^k(C_g^k + A_g^k B_g^k) + \sum_{j=1}^{K} q_{kj}A_g^j - \delta_0^k A_g^k \\
0 &= -\frac{dC_g^k}{d\tau} - \left(\kappa_g^k - \kappa_g^k\theta_g^k B_g^k - (\sigma_g^k)^2 B_g^k + \delta_0^k\right) C_g^k \\
&\quad - A_g^k\left(\frac{dB_g^k}{d\tau} + \kappa_g^k B_g^k - \frac{1}{2}(\sigma_g^k)^2(B_g^k)^2 + \delta^k\right) + \sum_{j=1}^{K} q_{kj}[C_g^j + A_g^j(B_g^j - B_g^k)] \\
0 &= -C_g^k\left(\frac{dB_g^k}{d\tau} + \kappa_g^k B_g^k - \frac{1}{2}(\sigma_g^k)^2(B_g^k)^2 + \delta^k\right) + \sum_{j=1}^{K} q_{kj}C_g^j(B_g^j - B_g^k),
\end{aligned} \tag{4.2.10}$$

with initial conditions

$$A_g^k(0, \mathbf{x}, \mathbf{y}) = x_k + y_k\delta_0^k, \ C_g^k(0, \mathbf{x}, \mathbf{y}) = y_k\delta^k, \ B_g^k(0, \mathbf{x}, \mathbf{y}) = 0. \tag{4.2.11}$$

where x_k, y_k are the k-th component of the vector $\mathbf{x}, \mathbf{y}$.

4.2.5 Pricing sovereign credit default swap spreads

A sovereign credit default swap (CDS) buyer pays a constant premium C for an insurance from the seller on the reference defaultable sovereign country. When the sovereign government bond defaults, the buyer gets a one-time cash settlements in the amount of L and stops paying any remaining premiums, where L is the loss rate of the defaultable bond.

To compute the value of the premium (fixed) leg of a CDS, we substitute into equation (4.2.5) the periodical payments of $P(T_m) = C\Delta T$, where T_m, $m = 1, \cdots, M$ are premium payment times to date for the on-the-run CDS, $\Delta T = T_m - T_{m-1}$ is the time interval between premium payments, and C is the nominal premium. The value of the premium leg becomes

$$\begin{aligned} P_{fx}(t,T) &= \sum_{m=1}^{M} C\Delta T \cdot P_0(t,t+T_m)\mathbb{E}_t\left[\exp\left(-\int_t^{t+T_m} z(a)da\right)\Phi(t,t+T_m)P(T_m)\right]\mathbf{1} \\ &= C\Delta T\sum_{m=1}^{M} P_0(t,t+T_m)\Lambda_z(T_m,z(t),1,0)\Lambda_g(T_m,g(t),\mathbf{1},0), \end{aligned}$$

For the default (floating) leg, substituting $P(T) = 0$ and $P^D(s) = \mathbf{L}$ into equation (4.2.5) yields

$$\begin{aligned} P_{fl}(t,T) &= \int_t^{t+T} P_0(t,s)\mathbb{E}_t\left[\exp\left(-\int_t^s z(a)da\right)\Phi(t,s)[G(s)+z(s)I]\mathbf{L}\right]ds \\ &= \int_0^T P_0(t,t+u)[\Lambda_z(u,z(t),1,0)\Lambda_g(u,g(t),0,\mathbf{L}) + \Lambda_z(u,z(t),0,1)\Lambda_g(u,g(t),\mathbf{L},0)]du. \end{aligned}$$

Equaling the premium leg and default leg, we obtain the CDS pricing formula, given the current regime $R(t) = k$, as

$$\begin{aligned} C(t,g^k(t),z(t)) &= \frac{\mathbf{1}_k^T P_{fl}(t,T)}{\mathbf{1}_k^T\sum_{m=1}^{M} P_0(t,t+T_m)\mathbb{E}_t\left[\exp\left(-\int_t^T z(a)da\right)\Phi(t,T_m)P(T_m)\right]\Delta T} \\ &= \frac{\int_0^\tau P_0(t,t+u)[\Lambda_z(u,z(t),1,0)\Lambda_g^k(u,g(t),0,L) + \Lambda_z(u,z(t),0,1)\Lambda_g^k(u,g(t),L,0)]du}{\sum_{m=1}^{M} P_0(t,t+T_m)\Lambda_z(T_m,z(t),1,0)\Lambda_g^k(T_m,g(t),\mathbf{1},0)\Delta T} \end{aligned} \tag{4.2.12}$$

4.2.6 Market price of risk and physical transition probability

In the pricing equations above, we use the risk-neutral measure and take the expectations with respect to the distribution associated with a hypothetical investor who is neutral towards the risk of unpredictable variation in the risk factors. To estimate the parameters of the risk factors from market data, however, we will need the physical measure and the probability distribution implied by the historical data. We use the superscript $\mathbb{P}$ to denote the parameters of the risk factors under the physical measure, and use the superscript $\mathbb{P}^*$ or no superscript to denote the parameters under the risk-neutral measure.

Under the physical measure $\mathbb{P}$, the risk factors $g(t), z(t)$ are assumed to also follow the mean-reverting process,

$$\begin{aligned} dz(t) &= \kappa_z^{\mathbb{P}}[\theta_z^{\mathbb{P}} - z(t)]dt + \sigma_z dW_z^{\mathbb{P}}(t) \\ dg(t) &= \kappa_g^{k,\mathbb{P}}[\theta_g^{k,\mathbb{P}} - g(t)]dt + \sigma_g^k dW_g^{\mathbb{P}}(t), \quad R(t-) = k. \end{aligned}$$

The dynamics of the risk factors under these two probability measures $\mathbb{P}$ and $\mathbb{P}^*$ are connected by the "market price of risk"

$$\eta(t) = \begin{pmatrix} \sigma_g^k\sqrt{g(t)} & 0 \\ 0 & \sigma_z\sqrt{z(t)} \end{pmatrix}^{-1} \left\{ \begin{pmatrix} \lambda_{g,0}^k \\ \lambda_{z,0} \end{pmatrix} + \begin{pmatrix} \lambda_{g,1}^k & 0 \\ 0 & \lambda_{z,1} \end{pmatrix} \begin{pmatrix} g(t) \\ z(t) \end{pmatrix} \right\}, \qquad (4.2.13)$$

at regime k. The change of probability distribution from risk-neutral measure $\mathbb{P}^*$ to historical measure $\mathbb{P}$ implies that the parameters satisfy

$$\begin{aligned} \kappa_z^{\mathbb{P}} &= \kappa_z - \lambda_{z,1}, \\ \theta_z^{\mathbb{P}}\kappa_z^{\mathbb{P}} &= \kappa_z\theta_z + \lambda_{z,0}, \\ \kappa_g^{k,\mathbb{P}} &= \kappa_g^k - \lambda_{g,1}^k, \\ \theta_g^{k,\mathbb{P}}\kappa_g^{k,\mathbb{P}} &= \kappa_g^k\theta_g^k + \lambda_{g,0}^k. \end{aligned}$$

This market price of risk allows the reverting speed and average level of the risk factors to differ across the physical measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{P}^*$, while assuring the factors $g(t), z(t)$ follow square-root mean-reverting processes under both measures.

We also assume the price of risk associated with the regime-shifts be

$$\Gamma = \begin{pmatrix} \Gamma_{11} & 0 \\ 0 & \Gamma_{22} \end{pmatrix}.$$

Thus, the transition matrix under the physical measure $\mathbb{P}$ is

$$Q^{\mathbb{P}} = \Gamma Q = \begin{pmatrix} q_{11}\Gamma_{11} & -q_{11}\Gamma_{11} \\ -q_{22}\Gamma_{22} & q_{22}\Gamma_{22} \end{pmatrix}$$

4.3 Maximum Likelihood Estimation

Different methods can be used to estimate model parameters in a regime-switching model. Gibbs sampling was developed for regime switching models by Albert and Chib [2] and Kim and Nelson [66]. Alternatively, the maximum likelihood method has gained much popularity since Hamilton [57], or in more details, Dai et al. [33]. The maximum likelihood algorithm contains a Bayesian updating procedure that computes the probability of being in a regime given all available information up to that time.

Given our model structure, we now proceed with the maximum likelihood estimation for the regime-switching models. Similar to Dai et al. [33], we assume that the five-year CDS spreads are priced exactly, so that the default intensity can be inferred by inverting the pricing function. The CDS spreads with each of the other maturities are assumed to be priced with normally distributed errors with mean zero and constant variance.

We assume two regimes $R = \{1, 2\}$. Let $\hat{c}_t$ be the CDS spreads priced exactly by the model, then in regime $s_t = j$,

$$\hat{c}_t = C(t, g_t^j, z_t), \tag{4.3.1}$$

and we can obtain g_t^j and z_t [1] by solving the above equations.

The remaining spreads used in estimation are denoted by $\tilde{c}_t$, with price error u_t^j assumed to be i.i.d Gaussian with mean zero and variance Ω^j in regime $s_t = j$. The information set $J_t = \{\hat{c}_\tau, \tilde{c}_\tau, \tau \leq t\}$ contains the observed price history up to time t, and $S_t^j = \mathbb{P}(s_t = j|J_t)$ is the probability of regime j given the current information.

Given the assumptions above, we can construct the likelihood function under physical measure $\mathbb{P}$ for the observed CDS spreads

$$\begin{aligned} f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t) &= \sum_{j,k} f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t, s_t = j, s_{t+1} = k) f(s_t = j, s_{t+1} = k|J_t) \\ &= \sum_{j,k} f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t, s_t = j, s_{t+1} = k) f(s_{t+1} = k|s_t = j, J_t) f(s_t = j|J_t) \\ &= \sum_{j,k} f(\tilde{c}_{t+1}|\hat{c}_{t+1}, J_t, s_t = j, s_{t+1} = k) f(\hat{c}_{t+1}|J_t, s_t = j, s_{t+1} = k) \pi_{jk}^{\mathbb{P}} S_t^j \end{aligned} \tag{4.3.2}$$

The Log-likelihood is then

$$\log L = \frac{1}{M-1} \sum_{t=0}^{M-1} \log f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t).$$

To complete the construction in (4.3.2), we need two terms: *the first term*, $f(\tilde{c}_{t+1}|\hat{c}_{t+1}, J_t, s_t = j, s_{t+1} = k)$, calculates the likelihood of the spreads priced with errors, given the perfectly priced spreads ; *the second term*, $f(\hat{c}_{t+1}|J_t, s_t = j, s_{t+1} = k)$, updates the likelihood of next observation of the perfectly priced spreads given current information.

[1] For convenience, we denote $g_t, t = 1, 2, \cdots$ to be the discrete sample of $g(t\Delta t)$ and similarly for c_t, z_t, s_t. Here Δt is the time increment of observation, for example, $\Delta t = 5/255$ for weekly observation of market data

We first derive the formula for *the first term*. When $s_{t+1} = k$,

$$\tilde{c}_{t+1} = C(t + \Delta t, g^k_{t+1}, z_{t+1}) + u^k_{t+1}, \quad u^k_{t+1} \sim N(0, \Omega^k)$$

where g^k_{t+1}, z^k_{t+1} are functions of $\hat{c}_{t+1}$, which are numerically solved using the pricing equation $\hat{c}_{t+1} = C(t + \Delta t, g^k_{t+_1}, z_{t+1})$. Thus the conditional density is

$$\begin{aligned} &f(\tilde{c}_{t+1}|\hat{c}_{t+1}, J_t, s_t = j, s_{t+1} = k) = f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_t = j, s_{t+1} = k) \\ &= \frac{\exp\left\{-\frac{1}{2}\left[\tilde{c}_{t+1} - C(t + \Delta t, g^k_{t+1}, z_{t+1})\right]' |\Omega^k \Omega^{k\prime}|^{-1} \left[\tilde{c}_{t+1} - C(t + \Delta t, g^k_{t+1}, z_{t+1})\right]\right\}}{\sqrt{2\pi|\Omega^k \Omega^{k\prime}|}} \end{aligned} \tag{4.3.3}$$

We then derive *the second term*. Under the physical measure $\mathbb{P}$, denote $f_z(x|z_t)$ be the conditional density function of z_{t+1} given z_t. Similarly, denote by $f_g(x|g^j_t)$ the conditional density function of $g_{t+1} = x$ given g^j_t in regime j. The second term is

$$\begin{aligned} f(\hat{c}_{t+1}|J_t, s_t = j, s_{t+1} = k) &= f(\hat{c}_{t+1}|\hat{c}_t, s_t = j, s_{t+1} = k) \\ &= f(\hat{c}_{t+1}|g^j_t, z_t; s_t = j, s_{t+1} = k) \\ &= f(C(t + \Delta t, g^k_{t+1}, z_{t+1})|g^j_t, z_t; s_t = j, s_{t+1} = k) \\ &= f_g(g^k_{t+1}|g^j_t) f_z(z_{t+1}|z_t)[\det(\partial C(t + \Delta t, g, z))]^{-1}|_{(g,z)=(g^k_{t+1}, z_{t+1})}. \end{aligned} \tag{4.3.4}$$

where $\partial C(t + \Delta t, g, z)$ is the Jacobian matrix of the pricing function C with respect to g and z. The last equality in the above formula is obtained by transformation formula of random variable.

Since $z(t)$ and $g(t)$ are independent square-root mean-reverting processes, the conditional distributions are known. Specifically, the value z_{t+1}, conditioned on z_t, follows a scaled non-central chi-squared distribution. Denote $\chi^2_{\text{pdf}}(x, s, d)$ as the density function of the non-central chi-squared distribution with scale parameter s and freedom d, we have the conditional density function of z_{t+1} given z_t as

$$f_z(x|z_t) = c_z \cdot \chi^2_{\text{pdf}}(xc_z, e^{-\kappa_z \Delta t} z_t c_z, d_z)$$

where the scale $c_z = 4\kappa_z^{\mathbb{P}}/[(\sigma_z)^2(1-e^{-\kappa_z^{\mathbb{P}}\Delta t})]$ and freedom parameter $d_z = 4\kappa_z^{\mathbb{P}}\theta_z^{\mathbb{P}}/(\sigma_z)^2$. Similarly, conditional on g_t^j and regimes $s_t = j, s_{t+1} = k$, we have

$$f_g(x|g_t^j) = c \cdot \chi^2_{\text{pdf}}(xc, e^{-\kappa_g^{k,\mathbb{P}}\Delta t} g_t^j c, d),$$

with $c = 4\kappa_g^{k,\mathbb{P}}/[(\sigma_g^k)^2(1-e^{-\kappa_g^{k,\mathbb{P}}\Delta t})]$ and $d = 4\kappa_g^{k,\mathbb{P}}\theta_g^{k,\mathbb{P}}/(\sigma_g^k)^2$.

The calculation of the non-central chi-squared density involves the evaluation of Bessel functions of the first kind, which can be numerically burdensome. Instead, we use Yacince's method (see Table 2 in Aït-Sahalia [1]) to get a closed-from approximation to the CIR density. This method was proved to have high-order accuracy and fast implementation.

The regime probability S_{t+1}^k is updated recursively using Bayes' rule

$$\begin{aligned}
&S_{t+1}^k = f(s_{t+1} = k|J_{t+1}) \\
&= f(s_{t+1} = k|\hat{c}_{t+1}, \tilde{c}_{t+1}, J_t) \\
&= \frac{\sum_j f(\hat{c}_{t+1}, \tilde{c}_{t+1}, s_{t+1} = k|J_t, s_t = j) S_t^j}{f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t)} \\
&= \frac{\sum_j f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_{t+1} = k, s_t = j, J_t) f(\hat{c}_{t+1}|s_{t+1} = k, s_t = j, J_t) f(s_{t+1} = k|s_t = j, J_t) S_t^j}{f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t)} \\
&= \frac{\sum_j f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_{t+1} = k, s_t = j) f(\hat{c}_{t+1}|\hat{c}_t, s_t = j, s_{t+1} = k) \pi_{jk}^{\mathbb{P}} S_t^j}{f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t)}
\end{aligned} \tag{4.3.5}$$

where the terms in the summation are given by (4.3.3) and (4.3.4), and the denominator is given by (4.3.2).

Remark 4.3.1. Note the parameters used in CDS pricing equations are those under the risk-neutral measure $\mathbb{P}^*$, and those used in likelihood function based on observation are under the physical measure $\mathbb{P}$.

Using matrix notation, we denote

$$S_t = \begin{pmatrix} S_t^1 & S_t^2 \end{pmatrix}, \tag{4.3.6}$$

$$f_{t,t+1}^c = \begin{pmatrix} f(\hat{c}_{t+1}|\hat{c}_t, s_t = 1, s_{t+1} = 1) & f(\hat{c}_{t+1}|\hat{c}_t, s_t = 1, s_{t+1} = 2) \\ f(\hat{c}_{t+1}|\hat{c}_t, s_t = 2, s_{t+1} = 1) & f(\hat{c}_{t+1}|\hat{c}_t, s_t = 2, s_{t+1} = 2) \end{pmatrix}, \tag{4.3.7}$$

$$f_{t,t+1}^u = \begin{pmatrix} f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_t = 1, s_{t+1} = 1) & f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_t = 1, s_{t+1} = 2) \\ f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_t = 2, s_{t+1} = 1) & f(\tilde{c}_{t+1}|\hat{c}_{t+1}, s_t = 2, s_{t+1} = 2) \end{pmatrix} \tag{4.3.8}$$

$$\hat{\pi} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 1 - \pi_1 & \pi_1 \\ \pi_2 & 1 - \pi_2, \end{pmatrix} \tag{4.3.9}$$

The log-likelihood is given by

$$\log L = \frac{1}{M-1} \sum_{t=0}^{M-1} \log f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t), \tag{4.3.10}$$

$$f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t) = S_t \times (f_{t,t+1}^c \odot f_{t,t+1}^u \odot \hat{\pi}) \times \mathbf{1}, \tag{4.3.11}$$

$$S_{t+1} = \frac{S_t \times (f_{t,t+1}^c \odot f_{t,t+1}^u \odot \hat{\pi})}{f(\hat{c}_{t+1}, \tilde{c}_{t+1}|J_t)} \tag{4.3.12}$$

where $\odot$ denotes element-by-element multiplication of matrices, $\times$ denotes matrix multiplication, and $\mathbf{1}$ is a unit vector with one row and two columns.

To interpret the regimes from the probabilities S_t^i, we use the "smoothed regime probabilities" $p_t^j = f(s_t = j|J_t)$ to classify the observations. If $p_t^j > 0.5$, we classify the observation at date t to regime j. The probability p_t^j can be computed by

$$p_t^j = \frac{I_t^j S_t^j}{\sum_k I_t^k S_t^k}$$

where I_t^j is calculated backwards in time, with $I_T^j = \mathbf{1}$ and

$$\begin{aligned} I_t^j &= f(\tilde{c}_l, \bar{c}_l : t+1 < l < T|s_t = j, J_t) \\ &= \sum_k \pi_{jk}^{\mathbb{P}} f(\tilde{c}_l, \bar{c}_l|J_t, s_t = j, s_{t+1} = k) I_{t+1}^k \end{aligned}$$

In matrix notation,

$$
\begin{aligned}
p_t &= \begin{pmatrix} p_t^1 \\ \\ p_t^2 \end{pmatrix} = \frac{S_t' \odot I_t}{S_t I_t}, \quad 1 \le t \le T \\
I_T &= \begin{pmatrix} 1 \\ \\ 1 \end{pmatrix}, \\
I_t &= \begin{pmatrix} I_t^1 \\ \\ I_t^2 \end{pmatrix} = (\pi^{\mathbb{P}} \odot f^c_{t,t+1} \odot f^u_{t,t+1}) \times I_{t+1}, \quad 1 \le t \le T-1
\end{aligned}
$$

4.3.1 Estimation of the global risk factor and country-specific factors

Note that in the pricing model, g_t is a global factor applying to all countries. If we perform the likelihood estimation country by country, we will most likely end up with different g_t's. If we perform the estimation with data from all of the sovereign counties together to get global factor g_t and country-specific factor z_t, the scale of computing is too large to be practical. Therefore, we perform the computation in two steps. The first step is to extract and estimate the global factor g_t and its parameters from credit default swaps of the United States by assuming that $z_t^{U.S.} \equiv 0$.

The likelihood function is the same as described in equation (4.3.2), except for some appropriate modifications in (4.3.3) and in (4.3.4). In (4.3.3), z_{t+1} takes the value of 0. In (4.3.4), function f_z vanished, and the Jacobian formula is written with respect to only g. That is, equation (4.3.4) in the likelihood estimation becomes

$$
f(\hat{c}_{t+1}|J_t, s_t = j, s_{t+1} = k) = f_g(g_{t+1}^k | g_t^j) \cdot [\det(\partial C(t + \Delta t, g, 0))]^{-1}|_{g=g_{t+1}^k}. \tag{4.3.13}
$$

The second step is to perform an estimation of the z_t's and their parameters for each of the sovereign country individually.

The five-year spread is assumed to be priced without error, and spreads with other maturities are assumed to be priced with Gaussian errors with mean zero. The

five-year spreads are chosen as benchmark since they are more liquid and actively traded in larger volume than those spreads with other maturities in general. Note that in the pricing equation (4.3.1) we take g_t as given, and the information set becomes $J_t = \{\hat{c}_\tau, \tilde{c}_\tau, g_\tau, Q^j_\tau, \tau \leq t, j = 1, 2\}$. The likelihood functions (4.3.2) and (4.3.3) remain unchanged and the Jacobian formula (4.3.4) involves only z. That is, equation (4.3.4) becomes

$$f(\hat{c}_{t+1}|J_t, s_t = j, s_{t+1} = k) = f_z(z_{t+1}|z_t) \cdot [\det(\partial C(t + \Delta t, g, z))]^{-1}|_{z=z^k_{t+1}}. \quad (4.3.14)$$

4.4 Empirical Results

In this section we begin by briefly describing the data used in the empirical analysis. We then describe the procedure of maximum likelihood estimation and discuss the empirical results obtained from the estimation.

4.4.1 Data

We use weekly CDS spreads with quarterly premium payments quoted in basis points,[2] where the CDS spreads in the United States are available from December 11, 2007 to February 15, 2011. The other sovereign countries include Brazil, Korea, Mexico, Russia, and South Africa. For each country, time series are available in full-time structure for seven fixed maturities: one, two, three, four, five, seven and ten years. The zero-coupon bond prices are extracted from Libor rates and swap rates by bootstrapping, the details of which are listed in Appendix 4.5.2. By matching the common time period of both CDS spreads and Libor rates data, we obtain 159 available observations.[3]

Table 4.1 provides summary statistics for the CDS spreads in the U.S. and the five

[2] One basis point is equivalent to 0.01%.

[3] There are 749 observations available for daily spreads. However we found that using weekly spreads seems to produce more stable results and the estimation is faster.

foreign countries. The top section shows the average levels of CDS spreads for each maturity in each country. The average values of the premiums vary widely across different countries, from a low of 34.90 basis points for 5-year spreads of the U.S to a high of 250.06 basis points for those of Russian. This variety of CDS average values indicates heterogeneity across different countries. The middle section gives the standard deviations, which are of similar magnitude to the previous mentioned average levels, indicating a significant time-series variation in the spreads. In addition, the standard deviations do not necessarily decrease as the terms increase. This might suggest persistency of credit risk in the risk-neutral measure, since otherwise less variation over the longer term would have been expected if the risk factor followed a strong mean-reversion. The bottom section presents the auto-correlations in the time-series, which are close to one, ranging from 0.8757 to 0.9723. This may suggest that the CDS spreads, and hence the credit risk factors, are persistent under statistical measure $\mathbb{P}$.

Figure 4.1 plots the time series of CDS spreads at selected maturities of one year (dashed blue lines), five years (solid green lines), and ten years (dash-dotted red lines). Each panel represents one country. The six chosen countries all experienced dramatic changes in CDS spreads during the sample period, especially during the sub-prime financial crisis. The three curves in each panel also show the variation of CDS term structure. The spreads with a longer term are in general wider than those with a shorter term, especially during normal time periods. The term structure can, however, revert during a crisis period. For example, from Russian CDS time series in the last panel of Figure 4.1, we see that the one-year spreads are much higher than the five-year and ten-year spreads during the first two quarters of 2009. This inverted term-structure curve, with wider spreads for shorter maturity, suggests that

Russia underwent a period of distressed credit.

4.4.2 Estimation

Several normalizations and restrictions are applied on the parameters as follows.

- To distinguish the two regimes, we let $\sigma_g^1 < \sigma_g^2$ and hence refer to regime 1 as "low-volatility regime" and regime 2 as "high-volatility regime." This restriction is intuitive, since we would expect a more volatile risk factor in the regime associate with more uncertainty. In addition, the estimation of volatility is generally more accurate than the estimation of means, as the mean can only be pinned down by long time series (see Merton [78]). Ang and Timmermann [9] tested estimations of a regime switching model on equity excess returns. They cannot reject the hypothesis that the means in two regimes are equal, but overwhelmingly reject the hypothesis that volatilities in two regimes are the same.

- To ensure that the mean-reverting risk factor $g(t)$ remains positive, we require $2\kappa_g^i\theta_g^i > \sigma_g^i$ for each regime $i \in \{1, 2\}$.

- To facilitate numerical identification of other free parameters, we assume that the sensitivity index does not depend on regimes, that is, $\delta_0^1 = \delta_0^2, \delta^1 = \delta^2 = \delta$. In addition, assume the loss rate $L = 40\%$.

With these normalization and constraints, the parameters to be estimated include

$$\kappa_g^1, \kappa_g^2, \theta_g^1, \theta_g^2, \sigma_g^1, \sigma_g^2, \delta_0, \delta, q_1, q_2, \lambda_{g,0}^1, \lambda_{g,0}^2, \lambda_{g,1}^1, \lambda_{g,1}^2, \Gamma_{11}, \Gamma_{22}, \Omega_{1,2,3,4,7,10y}.$$

Since the estimation involves a high dimensional parameter space as above, the choice of initial parameters is important. To get a fair initial guess, we follow the steps below.

- We first obtain an initial guess of $\kappa_g^i, \theta_g^i, \sigma_g^i$ using the five-year U.S CDS spreads. To do this, we split the data into two sets: one time-series before October 3, 2008 and one after. For each of the two time-series, we assume a single regime and zero market price of risks, perform the maximum likelihood estimation to find the mean-reversion parameters, then assign the result to low and high - volatility regimes respectively. The intuition behind this is that we expect that in the normal period before the sub-prime crisis the market is mainly in a low-volatility regime, while during and after the crisis, it switched to a high-volatility regime.

- Next we utilize the full-term structures by using the one-year up to ten-year U.S CDS spreads to estimate the regime switching parameters and the variances in pricing error. The market price of risks are still not considered in this step and will be estimated in the next step. Note that the mean-reversion parameters are also tuned to optimize the overall likelihood in this step.

- We then take the parameters found in the above steps as initial parameters and perform the full maximum likelihood estimation.

Once the dynamic of the global factor g_t and its parameters are obtained from the estimation in the first step, a second maximum likelihood estimation is performed on each sovereign country to obtain the country-specific factor z_t. For each of the seven countries, the parameters needed are

$$\kappa_z, \theta_z, \sigma_z, \delta_1, \lambda_{z,0}, \lambda_{z,1}, \Omega_{1y,2y,3y,4y,7y,10y}.$$

The maximum likelihood algorithms are implemented in MATLAB 2010a. To find a global maximizer of the likelihood, we use a direct pattern search method

until the result is close to a global optimum, then switch to a gradient-based or a derivative-free searching method to tune the results near the global optimum.

4.4.3 Empirical results

The time series of the estimated global factor g_t are shown in Figure 4.2. The systemic default risk factor increased significantly during the last quarter of 2008 after the bankruptcy of Lehman Brothers on September 15. It reached its peak of 277.7 basis points at the end of February 2009. It then decreased over the last three quarters of 2009. It rose again in November 2009 and reached a second peak near the end of February 2010, which may correlate with the large losses of Fannie Mae and public concerns about large deficits. After March 2010, it began to stabilize.

Figure 4.2 also presents the estimated probability of being in the high-volatility regime. According to National Bureau of Economic Research(NBER) business-cycle dating, the nearest period of recession starts from December 2007 to June 2009. Our model implies that the high-volatility regime starts from December 2007 to October 2008, shorter than the NBER defined recession period.

The estimated parameters for the global risk factor are presented in Table 4.2. We observed different behavior from the risk factor in the low-volatility v.s high-volatility regime with asymmetric regime-shift rates. The volatility of the global risk factor in the high-volatility regime is about twice as big as that in the low-volatility regime. The reverting speed κ, under both the physical P measure and the risk-neutral measure P^*, is much bigger in the high-volatility regime, suggesting that the risk factor is more persistent in the high-volatility regime than in the low-volatility regime.

The estimates of variance Ω_n measure the variance of the pricing errors for the CDS contracts with maturities of 1,2,3,4,7,10 years. They are typically small, on

the order of several basis points. For example, Ω_{2y} = 18.257E-08 implies that the standard deviations of the pricing error for one-year CDS spreads is 4.273 basis points . In addition, the standard deviation of pricing error tends to be smaller for intermediate maturities than short or long maturities. In other words, the model tends to fit the four-year spreads best with standard deviation of pricing error 1.212 basis points, and tends to fit the three-year and seven-year spreads (with standard deviation of pricing error of 2.913 and 2.618 respectively) better than one-year and ten-year spreads (with standard deviation of pricing error of 7.233 and 5.382 respectively). This make sense since we assumed the five-year spreads are priced without error and the rest of the spreads are priced with normally distributed error with mean zero and variance Ω_n.

Table 4.3 shows the estimated parameters for country-specific risk factor. Four out of five countries have a loading of systemic risk δ_1 greater than one. Brazil has biggest value of systemic index of 1.72, suggesting that it is most sensitive to the systemic risk. Russia, Brazil, South Africa and Mexico all have systemic index larger than one, indicating that their probability of default given a systemic shock is bigger than that of the United States. On the other hand, South Korea has δ_1 less than 1, implying that its default risk is more sovereign specific than systemic compared to other countries.

Brazil and Mexico exhibits larger volatility and faster reverting speed for country-specific risk factors than other countries. In terms of pricing error, Korea seems to show the best fit with relatively small Ω_n and large likelihood.

The time series of country-specific factors are presented in Figure 4.3. The sovereign-specific factors all experienced large increases in credit risk beginning with the sub-prime crisis in 2008, and reached their peak around September 2008 before

the systemic credit risk g_t reached its peak in 2009. The country-specific factor for Brazil showed a single high-rise and high-drop spike; Russia stayed at a high spread for a much longer time; while the other countries exhibit a v-shaped double dip during 2008 and 2009. The average magnitudes of sovereign-specific risks also vary across different countries.

4.4.4 Conclusion

This study investigated the properties of sovereign credit risk using credit default swap spreads for the United States and five major sovereign countries. Using a two-factor regime-switching credit model, we were able to decompose sovereign credit risk into a systemic risk factor and a country-specific risk factors. Maximum likelihood estimate was used to calibrate the model parameters to market data.

We discovered a high level of commonality in sovereign credit spreads and extracted the dynamics of the regime-switching systemic risk factor that contributes to this commonality. We also observed a large variation across different countries in terms of their exposure to systemic shocks and the dynamic of their country-specific risk factors.

4.5 Appendix

4.5.1 Derivation of the closed-form approximation for Λ_g

Plugging the expression (4.2.9) into the partial differential equation (4.2.8) and noticing the equalities

$$\begin{aligned}
\frac{\partial \Lambda_g^k}{\partial t} &= -(\frac{dA_g^k}{d\tau} + \frac{dC_g^k}{d\tau})e^{B_g^k g} - (A_g^k + C_g^k g)\frac{dB_g^k}{d\tau}e^{B_g^k g}, \\
\frac{\partial \Lambda_g^k}{\partial g} &= C_g^k e^{B_g^k g} + (A_g^k + C_g^k g)B_g^k e^{B_g^k g}, \\
\frac{\partial^2 \Lambda_g^k}{\partial g^2} &= 2C_g^k B_g^k e^{B_g^k g} + (A_g^k + C_g^k g)(B_g^k)^2 e^{B_g^k g},
\end{aligned}$$

and the first-order approximation

$$\begin{aligned}\{Q\Lambda_g\}^k &= \sum_{j=1}^{K} q_{kj}(A_g^j + C_g^j)e^{B_g^j g} \\ &= e^{B_g^k g}\sum_{j=1}^{K} q_{kj}(A_g^j + C_g^j)e^{(B_g^j - B_g^k)g} \\ &\approx e^{B_g^k g}\sum_{j=1}^{K} q_{kj}(A_g^j + C_g^j)[1 + (B_g^j - B_g^k)g],\end{aligned}$$

we can rewrite equation (4.2.8) explicitly as

$$\begin{aligned}0 = &-(\frac{dA_g^k}{d\tau} + \frac{dC_g^k}{d\tau}) - (A_g^k + C_g^k g)\frac{dB_g^k}{d\tau} \\ &+ \kappa_g^k(\theta_g^k - g)\left[C_g^k + [A_g^k + C_g^k g]B_g^k\right] + \frac{1}{2}g(\sigma_g^k)^2\left[2C_g^k B_g^k + [A_g^k + C_g^k g](B_g^k)^2\right] \\ &+ \sum_{j=1}^{K} q_{kj}(A_g^j + C_g^j)[1 + (B_g^j - B_g^k)g] - (\delta_0^k + \delta_1^k g)(A_g^k + C_g^k g).\end{aligned}$$

This equation holds for all g. By matching the terms, we obtain the ODE system (4.2.10) in Section 4.2.4.

4.5.2 Extracting the default-free zero-coupon bond prices from LIBOR and swap data

The default-free zero-coupon bond prices (discount factors) $P_0(t,T)$ for the CDS pricing equation in Section 4.2.5 are extracted from the U.S. LIBOR rates and swap rates [4] by bootstrapping with a standard liner interpolation algorithm.[5] The data were obtained from Bloomberg.

The available LIBOR rates $L(t,T)$ include those for overnight, one week, two week, and for one month up to twelve month. The discount factors based on these rates are

$$P(t,T) = \frac{1}{1 + TL(t,T)}. \tag{4.5.1}$$

[4] Swap rates are for a fixed rate payer in return for receiving three month LIBOR

[5] Alternative interpolation schemes are available, but the estimation results are not sensitive to the choice of the interpolation algorithm.

The swap rates are available for maturity of one year, two years, three years, etc , up to ten years. The semi-annual swap rate $s(T)$ with maturity T and the discount factor $P(t,k)$ satisfies

$$1 = \frac{s(T)}{2}\sum_{k=1}^{2T} P(t,\frac{k}{2}) + P(t,T).$$

With linear interpolation on the bond price such that $P(t,T)+P(t,T-1) = 2P(t,T-\frac{1}{2})$, we have

$$P(t,T) = \frac{1 - \frac{s(T)}{2}\sum_{k=1}^{2(T-1)} P(t,\frac{k}{2}) - \frac{s(T)}{4}P(0,T-1)}{1+\frac{3}{4}s(T)}. \tag{4.5.2}$$

Applying equation (4.5.1), equation (4.5.2) and linear interpolation, we can bootstrap the zero-curve $P(t,T)$ for every T on a given day.

4.5.3 The transition rate matrix and transition probability matrix

The stationary distribution for the continuous-time Markov Chain with rate matrix $Q = \begin{pmatrix} -q_1 & q_1 \\ q_2 & -q_2 \end{pmatrix}$ is given by $\left(\frac{q_2}{q_1+q_2}, \frac{q_1}{q_1+q_2}\right)$.

Noting Q can be diagonalized as

$$Q = A\begin{pmatrix} -(q_1+q_2) & 0 \\ 0 & 0 \end{pmatrix} A^{-1}, \tag{4.5.3}$$

where matrix $A = \begin{pmatrix} q_1 & 1 \\ -q_2 & 1 \end{pmatrix}$ and its inverse $A^{-1} = \frac{1}{q_1+q_2}\begin{pmatrix} 1 & -1 \\ q_2 & q_1 \end{pmatrix}$, we can obtain the one-step transition matrix as

$$\begin{aligned}(\pi)_{ij} = \exp(Q\Delta t) &= A\begin{pmatrix} e^{-(q_1+q_2)\Delta t} & 0 \\ 0 & 1 \end{pmatrix} A^{-1} \\ &= \frac{1}{q_1+q_2}\begin{pmatrix} q_1 e^{-(q_1+q_2)\Delta t} + q_2 & -q_1 e^{-(q_1+q_2)\Delta t} + q_1 \\ -q_2 e^{-(q_1+q_2)\Delta t} + q_2 & q_2 e^{-(q_1+q_2)\Delta t} + q_1 \end{pmatrix}\end{aligned} \tag{4.5.4}$$

Term		1 yr	2 yr	3 yr	4 yr	5 yr	7 yr	10 yr
Mean	U.S.	21.5782	28.0311	30.6606	33.2165	34.9040	37.3985	39.4581
	Brazil	91.1677	116.4472	134.1498	152.4452	165.8267	180.4143	194.6439
	Russia	210.3797	231.1624	241.8788	247.1236	250.9829	255.0645	259.7290
	Mexico	101.0520	122.7155	137.6096	155.1387	167.2919	180.1023	189.3149
	South Africa	119.3359	145.4896	167.0282	183.9909	195.9781	206.3323	213.5603
	South Korea	115.5827	125.2705	133.4531	141.4076	148.2501	154.4032	158.6413
Standard Deviation	U.S.	11.9447	16.1281	17.5830	18.8836	19.5185	19.8381	20.2655
	Brazil	73.1577	81.6996	84.5838	86.8887	88.4120	86.6050	84.0495
	Russia	296.7372	262.2709	238.1338	217.2570	204.3764	190.2685	183.0235
	Mexico	85.5295	91.8631	93.2407	94.6789	95.2834	94.3345	93.6841
	South Africa	116.2586	112.9740	109.1549	105.7984	103.4776	95.9319	90.4837
	South Korea	109.0316	110.8867	109.5936	105.9472	102.9308	98.7574	96.0723
Auto-Correlation	U.S.	0.9709	0.9715	0.9717	0.9715	0.9702	0.9698	0.9671
	Brazil	0.9056	0.9088	0.8993	0.8984	0.8911	0.8885	0.8831
	Russia	0.9578	0.9460	0.9363	0.9238	0.9126	0.8968	0.8867
	Mexico	0.9494	0.9358	0.9201	0.9137	0.9039	0.8979	0.8929
	South Africa	0.9366	0.9267	0.9127	0.9032	0.8937	0.8832	0.8757
	South Korea	0.9121	0.9106	0.9052	0.8988	0.8918	0.8892	0.8870

Table 4.1: Summary statistics of CDS spreads (in basis points).

		low-vol regime	high-vol regime
under $P*$	κ_g^j	0.097735	1.073436
	θ_g^j	0.020842	0.000450
	σ_g^j	0.066532	0.184256
	q_{ij}	0.000092	0.274053
under P	κ_g^j	0.097735	1.073419
	θ_g^j	0.022822	0.016368
	σ_g^j	0.066532	0.184256
	q_{ij}	0.000058	0.175485
	$\lambda_{g,0}^j$	0.000193	0.017086
	$\lambda_{g,1}^j$	5.1829E-06	0.000018
	$\begin{pmatrix} \Gamma_{11} & 0 \\ 0 & \Gamma_{22} \end{pmatrix}$	$\begin{pmatrix} 0.631982 & 0 \\ 0 & 0.640347 \end{pmatrix}$	
	$\Omega_{(1y)}$	50.3217E-08	
	$\Omega_{(2y)}$	18.2569E-08	
	$\Omega_{(3y)}$	8.48761E-08	
	$\Omega_{(4y)}$	1.4678E-08	
	$\Omega_{(7y)}$	6.8558E-08	
	$\Omega_{(10y)}$	28.9663E-08	

Table 4.2: MLE estimates for global factor parameters using weekly U.S. CDS data. Maximized log-likelihood is 45.0523.

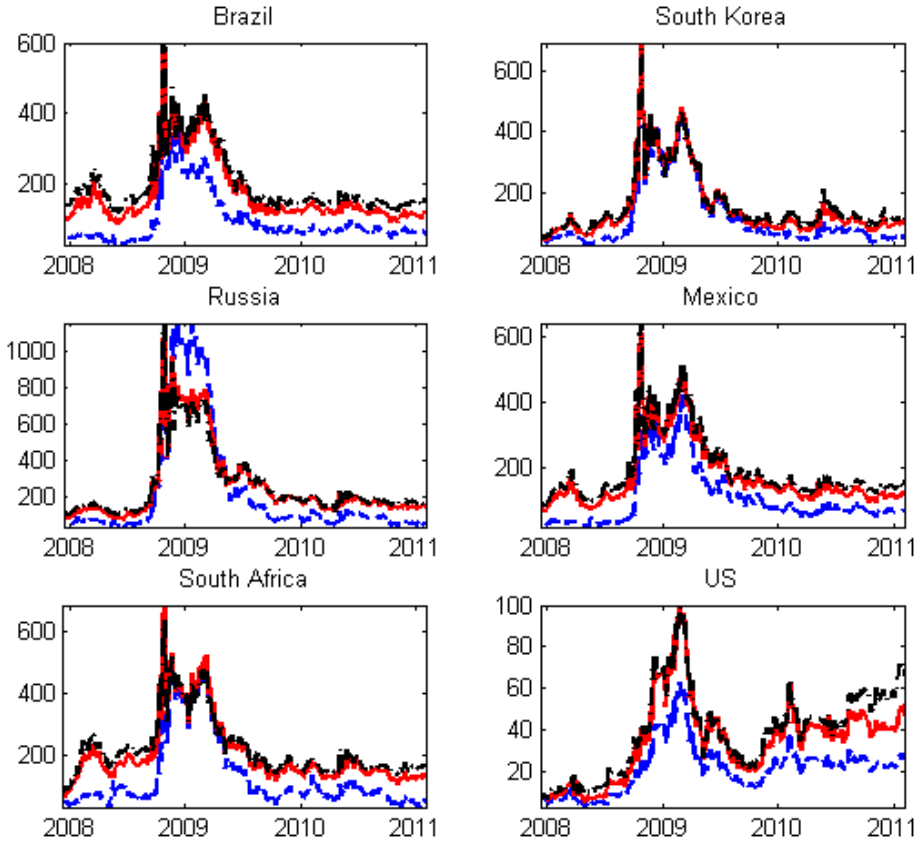


Figure 4.1: The time series of CDS spreads at selected maturities of one year (dashed blue lines), five years (solid red lines), and 10 years (dash-dotted black lines). Each panel is for one country.

	Brazil	Russia	South Africa	South Korea	Mexico
$\kappa_z^{\mathbb{P}}$	1.708736	0.189423	0.697728	0.034730	1.623650
$\theta_z^{\mathbb{P}}$	0.127030	0.100966	0.020892	0.399726	0.143155
$\sigma_z^{\mathbb{P}}$	0.641032	0.195578	0.170743	0.166627	0.619225
δ_1	1.716005	1.188747	1.184653	0.254693	1.104975
$\lambda_{0,z}$	0.209272	0.008130	0.000207	0.008867	0.203402
$\lambda_{1,z}$	0.005387	0.000014	0.000004	0.000002	0.000001
Ω_{1y}	0.00102960	0.00008359	0.00026426	0.00000417	0.00070749
Ω_{2y}	0.00029122	0.00002203	0.00012914	0.00000194	0.00021050
Ω_{3y}	0.00007872	0.00000726	0.00004098	0.00000118	0.00005909
Ω_{4y}	0.00001286	0.00000106	0.00000758	0.00000025	0.00000917
Ω_{7y}	0.00001342	0.00000179	0.00000791	0.00000042	0.00001017
Ω_{10y}	0.00004170	0.00000842	0.00002301	0.00000186	0.00002947
log-likelihood	24.8147	30.4090	25.9331	37.2523	25.4528

Table 4.3: Estimated parameters of the sovereign-specific risk factors z_t for the five countries. The parameters are calibrated by maximum likelihood estimation to weekly CDS spreads.

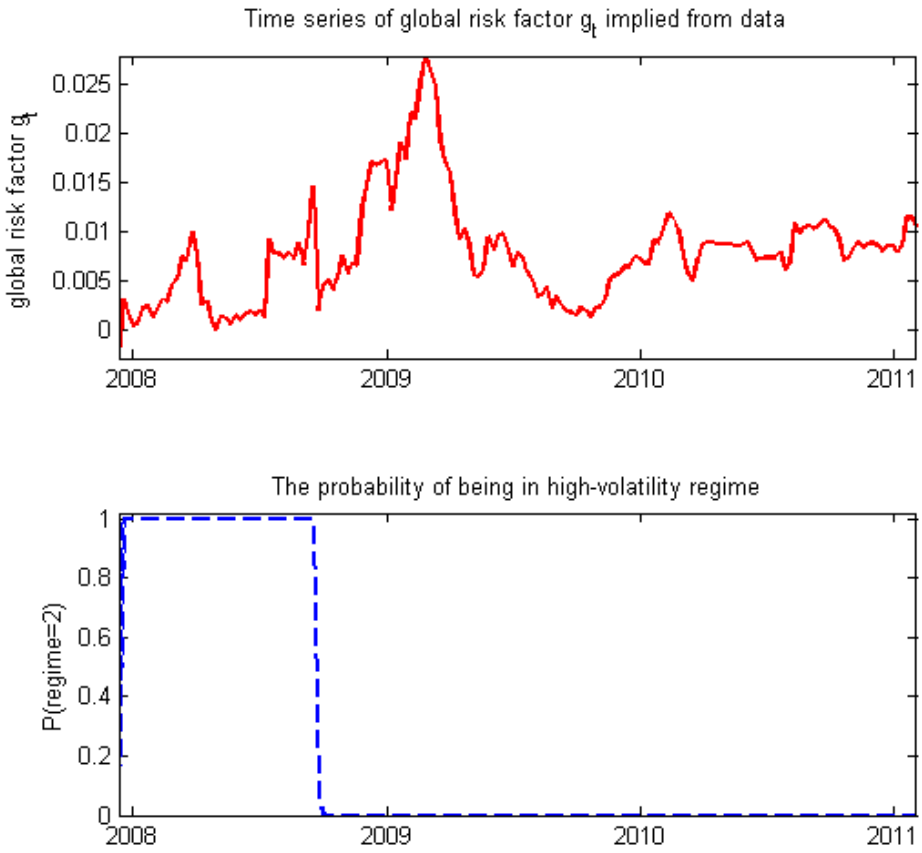


Figure 4.2: Estimated time series of global risk factor g_t (red solid curve above) and the probability of being in the high-volatility regime (blue dotted line below).

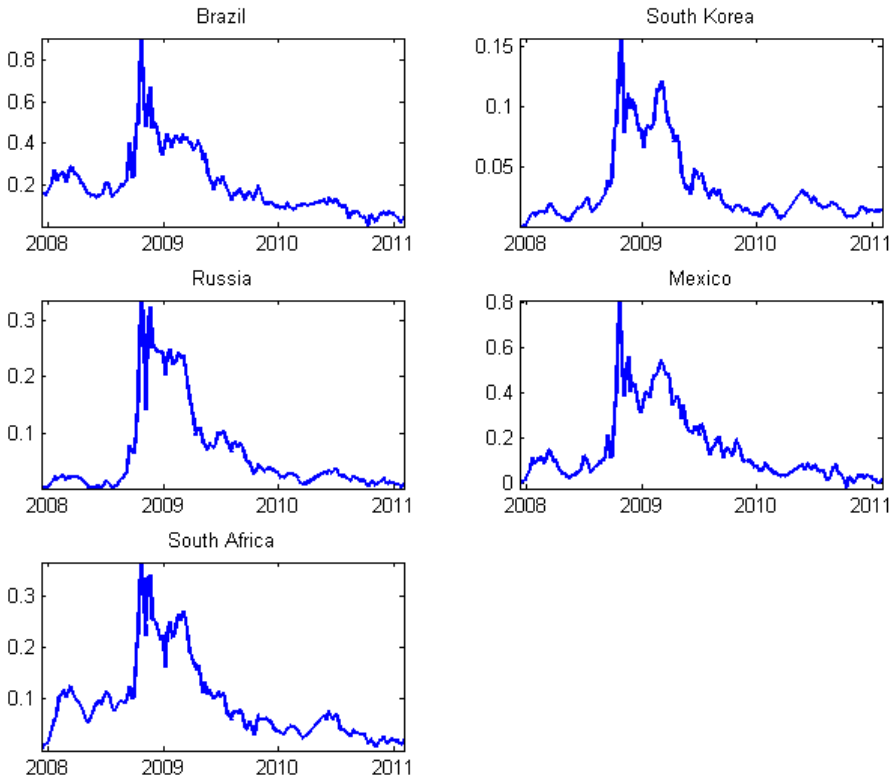


Figure 4.3: The time series of sovereign-specific factor z_t. Each panel represents one country.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *The Journal of Finance*, 54(4):1361–1395.

[2] 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 & Economic Statistics*, 11(1):1–15.

[3] Alfonsi, A. (2010). High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Mathematics of Computation*, 79(269):209–237.

[4] Andersen, L. (2008). Efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance*, 11(3):1–42.

[5] Ang, A. and Bekaert, G. (1998). Regime switches in interest rates. Technical report, National Bureau of Economic Research.

[6] Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial studies*, 15(4):1137–1187.

[7] Ang, A. and Bekaert, G. (2004). How regimes affect asset allocation. *Financial Analysts Journal*, 60(2):86–99.

[8] Ang, A. and Longstaff, F. (2011). Systemic sovereign credit risk: Lessons from the us and europe. Technical report, National Bureau of Economic Research.

[9] Ang, A. and Timmermann, A. (2011). Regime changes and financial markets. Technical report, National Bureau of Economic Research.

[10] Athreya, K. B., Kliemann, W., and Koch, G. (1988). On sequential construction of solutions of stochastic differential equations with jump terms. *Systems & Control Letters*, 10(2):141–146.

[11] Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049.

[12] Bally, V. and Talay, D. (1996). The law of the Euler scheme for stochastic differential equations. *Probability Theory and Related Fields*, 104(1):43–60.

[13] Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *Review of Financial Studies*, 9(1):69–107.

[14] Battles, Z. and Trefethen, L. (2004). An extension of matlab to continuous functions and operators. *SIAM Journal on Scientific Computing*, 25(5):1743–1770.

[15] Bayraktar, E., Giesecke, K., and Hu, X. (2011a). Exact simulation of Heston stochastic volatility model with jumps. *Working paper*.

[16] Bayraktar, E., Hu, X., and Young, V. (2011b). Minimizing the probability of lifetime ruin under stochastic volatility. *Insurance: Mathematics and Economics*, 49(2):43–78.

[17] Bayraktar, E., Moore, K. S., and Young, V. R. (2008). Minimizing the probability of lifetime ruin under random consumption. *North American Actuarial Journal*, 12(4):384–400.

[18] Bayraktar, E. and Young, V. (2011). Proving regularity of the minimal probabil-

ity of ruin via a game of stopping and control. *Finance and Stochastics*, 15(4):785–818.

[19] Bayraktar, E. and Young, V. R. (2007a). Correspondence between lifetime minimum wealth and utility of consumption. *Finance and Stochastics*, 11(2):213–236.

[20] Bayraktar, E. and Young, V. R. (2007b). Minimizing the probability of lifetime ruin under borrowing constraints. *Insurance: Mathematics and Economics*, 41(1):196–221.

[21] Bayraktar, E. and Young, V. R. (2008). Minimizing the probability of ruin when consumption is ratcheted. *North American Actuarial Journal*, 12(4):428–442.

[22] Bayraktar, E. and Young, V. R. (2009). Optimal deferred life annuities to minimize the probability of lifetime ruin. *North American Actuarial Journal*, 13(1):141–154.

[23] Benhamou, E., Gobet, E., and Miri, M. (2010). Time dependent Heston model. *SIAM Journal on Financial Mathematics*, 1:289–325.

[24] Berg, A. and Sachs, J. (1988). The debt crisis structural explanations of country performance. *Journal of Development Economics*, 29(3):271–306.

[25] Boehmer, E. and Megginson, W. (1990). Determinants of secondary market prices for developing country syndicated loans. *Journal of Finance*, 45(5):1517–1540.

[26] Boyd, J. P. (2001). *Chebyshev and Fourier spectral methods*. Dover Publications Inc., Mineola, NY, second edition.

[27] Broadie, M. and Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231.

[28] Carr, P. and Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73.

[29] Carr, P. and Schoutens, W. (2008). Hedging under the Heston model with jump-to-default. *International Journal of Theoretical and Applied Finance*, 11(4):1–12.

[30] Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341.

[31] Ceci, C. and Gerardi, A. (2006). A model for high-frequency data under partial information: a filtering approach. *International Journal of Theoretical and Applied Finance*, 9(4):555–576.

[32] Collin-Dufresne, P., Goldstein, R., and Martin, J. (2001). The determinants of credit spread changes. *The Journal of Finance*, 56(6):2177–2207.

[33] Dai, Q., Singleton, K. J., and Yang, W. (2007). Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields. *The Review of Financial Studies*, 20(5):1669–1706.

[34] Dooley, M. (2000). A model of crises in emerging markets. *The Economic Journal*, 110(460):256–272.

[35] Duffie, D. and Glynn, P. (1995). Efficient Monte Carlo simulation of security prices. *The Annals of Applied Probability*, 5(4):897–905.

[36] Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

[37] Duffie, D., Pedersen, L., and Singleton, K. (2003). Modeling sovereign yield spreads: a case study of Russian debt. *The Journal of Finance*, 58(1):119–159.

[38] Eichengreen, B. and Mody, A. (2000). What explains changing spreads on emerging-market debt: fundamentals or market sentiment? *The Economics of International Capital Flows*.

[39] Elices, A. (2009). Affine concatenation. *Wilmott Journal*, 1(3):155–162.

[40] Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *The Journal of Finance*, 59(3):1367–1404.

[41] Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *Journal of Finance*, 58(3):1269–1300.

[42] Farnsworth, H. and Li, T. (2007). The dynamics of credit spreads and ratings migrations. *Journal of Financial and Quantitative Analysis*, 42(03):595–620.

[43] Fouque, J.-P., Papanicolaou, G., and Sircar, K. R. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, Cambridge.

[44] Fouque, J.-P., Papanicolaou, G., Sircar, K. R., and Solna, K. (2003). Multiscale stochastic volatility asymptotics. *SIAM Journal on Multiscale Modeling and Simulation*, 2(1):22–42.

[45] Fouque, J.-P., Wignall, B., and Zhou, X. (2008). First passage model under stochastic volatility. *Journal of Computational Finance*, 11(3):43–78.

[46] Frey, R., Prosdocimi, C., and Runggaldier, W. J. (2007). Affine credit risk models under incomplete information. In *Stochastic processes and applications to mathematical finance*, pages 97–113. World Sci. Publ., Hackensack, NJ.

[47] Frey, R. and Runggaldier, W. (2010). Pricing credit derivatives under incomplete information: a nonlinear-filtering approach. *Finance and Stochastics*, 14(4):495–526.

[48] Geyer, A., Kossmeier, S., and Pichler, S. (2004). Measuring systematic risk in emu government yield spreads. *Review of Finance*, 8(2):171–197.

[49] Giesecke, K., Kakavand, H., and Mousavi, M. (2010). Exact simulation of point processes with stochastic intensities. *Operations Research, forthcoming.*

[50] Giesecke, K., Kim, B., and Zhu, S. (2011). Monte carlo algorithms for default timing problems. *Management Science*, (Articles in Advance):1–15.

[51] Giesecke, K. and Smelov, D. (2011). Exact sampling of jump-diffusions. *Operations Research, second round of review.*

[52] Glasserman, P. and Kim, K. (2011). Gamma expansion of the heston stochastic volatility model. *Finance and Stochastics*, 15(2):267–296.

[53] Gonzalez-Rozada, M. and Yeyati, E. (2008). Global factors and emerging market spreads. *The Economic Journal*, 118(533):1917–1936.

[54] Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

[55] Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544.

[56] Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies*, 21(2):889–935.

[57] Hamilton, J. (1988). Rational-expectations econometric analysis of changes in regime:: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12(2-3):385–423.

[58] Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2):357–384.

[59] Hamilton, J. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593.

[60] Hamilton, J. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2):307–333.

[61] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343.

[62] Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition.

[63] Jaeckel, P. and Kahl, C. (2006). Fast strong approximation Monte-Carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(06):513–536.

[64] Jonsson, M. and Sircar, K. R. (2002). Partial hedging in a stochastic volatility environment. *Mathematical Finance*, 12(4):375–409.

[65] Kamin, S. and Von Kleist, K. (1999). The evolution and determinants of emerging markets credit spreads in the 1990s. *BIS Working Paper No. 68*.

[66] Kim, C. and Nelson, C. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, vol. 1.

[67] Kliemann, W., Koch, G., and Marchetti, F. (1990). On the unnormalized solution of the filtering problem with counting process observations. *IEEE Transactions on Information Theory*, 36(6):1415–1425.

[68] Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin.

[69] Kushner, H. J. (1999). Consistency issues for numerical methods for variance control, with applications to optimization in finance. *IEEE Transactions on Automatic Control*, 44(12):2283–2292.

[70] Kushner, H. J. and Dupuis, P. (2001). *Numerical methods for stochastic control problems in continuous time*, volume 24. Springer Verlag, New York, second edition. Stochastic Modelling and Applied Probability.

[71] Leippold, M. and Wu, L. (2002). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis*, 37(2):271–295.

[72] Li, H., Li, T., and Hu, X. (2011). Pricing cds spreads with a regime switching model. *Working paper*.

[73] Li, H. and Xu, Y. (2009). Short rate dynamics and regime shifts. *International Review of Finance*, 9(3):211–241.

[74] Longstaff, F., Pan, J., Pedersen, L., and Singleton, K. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103.

[75] Lord, R., Koekkoek, R., and Van Dijk, D. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194.

[76] Mason, J. C. and Handscomb, D. C. (2003). *Chebyshev polynomials*. Chapman & Hall/CRC, Boca Raton, FL.

[77] Mauro, P., Sussman, N., and Yafeh, Y. (2002). Emerging market spreads: then versus now. *The Quarterly Journal of Economics*, 117(2):695–733.

[78] Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323 – 361.

[79] Meyer, P. A. (1971). Démonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités, V (Univ. Strasbourg, année universitaire 1969–1970)*, pages 191–195. Lecture Notes in Math., Vol. 191. Springer, Berlin.

[80] Mikhailov, S. and Nögel, U. (2003). Heston's stochastic volatility model implementation, calibration and some extensions. *Wilmott Magazine*, 4:74–79.

[81] Milevsky, M. A., Moore, K. S., and Young, V. R. (2006). Asset allocation and annuity-purchase strategies to minimize the probability of financial ruin. *Mathematical Finance*, 16(4):647–671.

[82] Moore, K. S. and Young, V. R. (2006). Optimal and simple, nearly optimal rules for minimizing the probability of financial ruin in retirement. *North American Actuarial Journal*, 10(4):145–161.

[83] Pachón, R. and Trefethen, L. N. (2009). Barycentric-Remez algorithms for best polynomial approximation in the chebfun system. *BIT*, 49(4):721–741.

[84] Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3–50.

[85] Pan, J. and Singleton, K. (2008). Default and recovery implicit in the term structure of sovereign cds spreads. *The Journal of Finance*, 63(5):2345–2384.

[86] Platen, E. and Bruti-Liberati, N. (2010). *Numerical solution of stochastic differential equations with jumps in finance*, volume 64 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin.

[87] Protter, P. E. (2004). *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition. Stochastic Modelling and Applied Probability.

[88] Remolona, E. M., Scatigna, M., and Wu, E. (2008). The dynamic pricing of sovereign risk in emerging markets:fundamentals and risk aversion. *The Journal of Fixed Income*, 17(4):57–71.

[89] Royston, P. and Altman, D. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, 43(3):429–467.

[90] Royston, P. and Altman, D. (1997). Approximating statistical functions by using fractional polynomial regression. *The Statistician*, 46(3):411–422.

[91] Schoutens, W. (2000). *Stochastic processes and orthogonal polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

[92] Scott, L. (1996). Simulating a multi-factor term structure model over relatively long discrete time periods. *Proceedings of the IAFE First Annual Computational Finance Conference*.

[93] Situ, R. (2005). *Theory of stochastic differential equations with jumps and appli-*

cations: mathematical and analytical techniques with applications to engineering. Springer Verlag.

[94] Smith, R. D. (2007). An almost exact simulation method for the Heston model. *Journal of Computational Finance*, 11(1):115–125.

[95] Todorov, V. and Tauchen, G. (2011). Volatility jumps. *Journal of Business and Economic Statistics*, 29(3):356–371.

[96] Van Haastrecht, A. and Pelsser, A. (2010). Efficient, almost exact simulation of the Heston stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 13(1):1–43.

[97] Young, V. R. (2004). Optimal investment strategy to minimize the probability of lifetime ruin. *North American Actuarial Journal*, 8(4):105–126.

[98] Zhang, F. X. (2003). What did the credit market expect of argentina default?: Evidence from default swap data. *Federal Reserve Board*, 3.