Statistical Inference and Pricing for Regime Switching Models in Finance and Insurance

by

Fangyuan Lin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis including any required final revisions as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis studies the estimation, goodness-of-fit testing, pricing and sampling problems for regime switching models, which are popularly used in financial markets. Specifically, we consider such models whose distributions are characterized by their characteristic functions, for example, Lévy processes. The thesis contains the following contents:

Chapter 1 introduces regime switching models and Lévy processes. Then we present the problems we would like to address in the following chapters and our main contributions to these problems.

Chapter 2 studies the estimation problem for regime switching Lévy processes. We extend an existing estimation method that is based on characteristic functions to our models. Meanwhile, we compare the estimation results obtained by the proposed estimation method with those obtained by the expectation-maximization (EM) algorithm. We also address several computational challenges within the proposed estimation method.

Chapter 3 studies the goodness-of-fit testing problem for regime switching models, where we extend two existing goodness-of-fit tests. Both of the proposed tests are based on characteristic functions.

Chapter 4 applies the estimation and testing methods proposed in Chapters 2 and 3 to a set of S&P 500 real data.

Chapter 5 studies the pricing problem for regime switching Lévy processes. We propose a numerical pricing method that provides a unified pricing framework. The proposed method is illustrated by pricing European and Bermudan options and ratchet equity-index annuities (EIAs) with surrender risk.

Chapter 6 studies the problem of sampling conditioned processes of regime switching models, where we propose an algorithm to sample paths from conditioned processes for a two-regime switching Black-Scholes model. Then we apply the proposed algorithm to the problems of pricing and static hedging of path-dependent options, where we use an Asian call option for illustrations.

Chapter 7 lists several topics for future research.

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Adam W. Kolkiewicz for the continuous support of my Ph.D studies and related research, for his patience, motivation, and immense knowledge. His intelligent ideas and insightful guidance helped me in all the time of research and writing of this thesis. He is also a good friend who gave me loads of valuable suggestions in my study, interview preparation and life in general. I could not have imagined having a better supervisor and mentor for my Ph.D studies.

I would also like to thank the rest of my thesis committee: Dr. Rogemar Mamon, Dr. Don McLeish, Dr. Tony Wirjanto and Dr. Ken Vetzal, for their insightful comments and encouragements.

I take this opportunity to express gratitude to all of the department faculty members for their help and support. They are Dr. Steve Brown, Dr. Johnny Li, Dr. Jock MacKay, Dr. Paul Marriott, Dr. Christopher Small, Dr. Ken Seng Tan, Dr. Ruodu Wang, Dr. Gord Willmot, Dr. Mu Zhu and many others. I also wish to specially thank Ms. Lisa Baxter, Ms. Leanne Bird, Ms. Mary Lou Dufton, Ms. Anthea Dunne and Ms. Karen Richardson for their help throughout my Ph.D. studies.

I am also grateful to acknowledge the financial support from the Department of Statistics and Actuarial Science, Waterloo Research Institute in Insurance, Securities and Quantitative Finance (WatRISQ) and the University of Waterloo.

Finally, my sincere thanks go to my colleagues, Lichen Chen, Dezhao Han, Feng He, Zhiyue Huang, Fangda Liu, Haiyan Liu, Kai Liu, Fei Meng, Xichen She, Jie Shen, Di Xu, Jinggong Zhang, Saisai Zhang, Wenjun Zhu and many others. We accompanied each other and had a wonderful time during those years.

Dedication

This is dedicated to my parents, Youhua Lin and Wenxia Fang, who always give emotional and physical care and support to me. Their unconditional and endless love and unhesitating and constant financial support motivate me to set clearer and higher goals in my study, work and life. This is also dedicated to my husband, Xin Yao, who always supports and encourages me whenever I face challenges in study and life. Thanks for your love and patience, and I am truly thankful for having you in my life.

Table of Contents

List of Tables

List of Figures

List of Notations

Here we provide a list of notations that are used in this thesis. Matrices and vectors are typically denoted by bold letters, for example, $\mathbf{z}$, and the transpose of $\mathbf{z}$ is denoted by $\mathbf{z}'$. For precise definitions and more abbreviations, see details in each chapter.

Symbol	**Description**
$\forall$	for all
$\exists$	there exists
$\cap$	the intersection of two measurable sets
P	the real-world measure
Q	a risk-neutral measure
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of positive real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers
$\mathbb{Z}^+$	the set of positive integers
$\Re(\cdot)$	real part of a complex number or function
$\Im(\cdot)$	imaginary part of a complex number or function
$\mathcal{W} = \{W_j, j \in \mathbb{Z}^+\}$	observable stochastic process, e.g., log-returns of the underlying asset
$\mathcal{Y} = \{Y_j, j \in \mathbb{Z}\}$	Markov chain with index j
$\mathbf{S} = \{k, k \in \mathbb{Z}^+\}$	state space of $\mathcal{Y}$
$\underline{\xi} = (\xi_1, \ldots, \xi_p)$	a vector of p model parameters
$Z(\underline{\xi}) = \{Z_{j+1}(\underline{\xi}), j \in \mathbb{Z}^+\}$	independent random variables that depend on $\underline{\xi}$, and they are assumed to be increments of a Lévy process in this thesis
$p_{kl}^{(n)}$	n-step transition probability of $\mathcal{Y}$
$\mathcal{P} = [p_{kl}, k, l \in \mathbf{S}]$	transition matrix of $\mathcal{Y}$
$\pi = \{\pi_k, k \in \mathbf{S}\}$	stationary distribution of $\mathcal{Y}$
$\mathcal{L} = \{L_t, t \geq 0\}$	a Lévy process
Φ_m	the m-dimensional characteristic function of the model
$\bar{\Phi}_{m,N}$	the m-dimensional ECF of the model with N observations
ζ_m	a vector of real and imaginary parts of Φ_m

$\zeta_{m,N}$	a vector of real and imaginary parts of $\bar{\Phi}_{m,N}$
Ω_m	the covariance matrix of $\zeta_{m,N}$
$\{\alpha_l, l \in \mathbb{Z}\}$	the mixing coefficients of an α-mixing process defined in Chapter 3
Q_N^i	simulated test statistic at i^{th} simulation with N observations
$Q_{\alpha,n}$	critical value for a χ^2 distribution with a given (right-tail) probability level α and the degrees of freedom n
h_e	the payoff function of the optimal hedging option that minimizes the mean-square of hedging error defined in Chapter 6
h_{opt}	the payoff function of the optimal hedging option that minimizes the mean-square of the shortfall risk defined in Chapter 6
$U(a, b)$	a uniform distribution on the interval $[a, b]$
$N(a, b)$	a normal distribution with mean a and variance b
$LN(a, b)$	a log-normal distribution with mean a and variance b
$POI(\lambda)$	a Poisson distribution with intensity λ
i.i.d.	independent and identically distributed
ECF	empirical characteristic function
CDF	cumulative distribution function
PDF	probability density function
RSBS	regime switching Black-Schoels
RSM	regime switching Merton
RSVG	regime switching variance gamma
DECF	discrete empirical characteristic function
FDAWO	finite difference approximation with optimization
GOF	goodness-of-fit
LR	likelihood ratio
PV	projected value
SMCP	sampling method for the conditioned process

Chapter 1

Introduction

Hidden Markov models (HMMs) (see, for example, Mamon and Elliott (2007)) are widely used in finance, signal processing, biology, psychology, geography, etc. In finance and economics, regime switching models (RSMs) represent one of the two main classes of applications of HMMs[1]. Although numerous authors have investigated the problems of estimation, goodness-of-fit testing, pricing and sampling for such models, few papers have discussed these problems for RSMs whose distributions are characterized by their characteristic functions only. In the thesis, we plan to address some of the outstanding issues for these models.

1.1 The Model

In this section, we will give the definitions and some examples of hidden Markov models whose characteristic functions can be written explicitly. All models are defined under the real-world P measure.

[1] As suggested by my committee member, Professor Tony Wirjanto, in the broadest sense, a HMM is a Markov process that has an observable (W_j) and unobservable (Y_j) components. In one class of applications, HMMs describe a setting where a stochastic system is observed through noisy measurements Y_j representing a random signal and W_j representing an observable corrupted version of the original signal. In the second class of applications, W_j itself is of interest while Y_j (such as an unobserved economic factor which induces fluctuations in stock prices) represents the influence on observable W_j (such as market prices of a stock). RSMs come from the second application of HMMs.

1.1.1 Regime Switching Models

Let $\mathcal{Y} := \{Y_j, j \in \mathbb{Z}\}$ be a Markov chain and $\mathbf{S}$ be its state space. Assume that $\mathbf{S}$ is finite, and its elements will be denoted by k, where $k \in \mathbb{Z}^+$. Let $\mathcal{P} = [p_{kl}]$ be the transition matrix and $p_{kl}^{(n)} = \Pr(Y_n = l \mid Y_0 = k)$ be the n-step transition probability of $\mathcal{Y}$, where $k, l \in \mathbf{S}$ and $n \in \mathbb{Z}^+$. If $n = 1$, then $p_{kl}^{(n)} = p_{kl}$. Define $\pi = \{\pi_k, k \in \mathbf{S}\}$ as a stationary distribution of $\mathcal{Y}$, that is, it satisfies $\sum_{k \in \mathbf{S}} \pi_k = 1$ and $\pi_l = \sum\limits_{k \in \mathbf{S}} \pi_k p_{kl}$. It is known (see, for example, Cinlar (2013)) that if $\mathcal{Y}$ is irreducible, aperiodic, and positive recurrent, then π is the unique stationary distribution of $\mathcal{Y}$, and for every $k, l \in \mathbf{S}$, $\lim\limits_{n \to \infty} p_{kl}^{(n)} = \pi_l$.

Next we define the hidden Markov models used throughout this thesis.

Let $\mathcal{Y} = \{Y_j, j = 0, ..., N-1\}$ be an unobservable Markov chain with K states and $\mathcal{W} = \{W_j, j = 1, ..., N\}$ be an observable stochastic process (for example, the log-returns of the underlying asset), where $N \in \mathbb{Z}^+$. We define a hidden Markov model as a bivariate process $(\mathcal{Y}, \mathcal{W})$ that satisfies the following conditions:

(A1–1) States/Regimes: $\mathcal{Y}$ can take values from the state space $\mathbf{S} = \{1, ..., K\}$.

(A1–2) Independence: $\mathcal{Y}$ is independent of the process $\mathcal{W}$; conditional on $\mathcal{Y}$, $\mathcal{W}$ is a sequence of independent variables.

(A1–3) Homogeneity: $\mathcal{Y}$ is a time-homogeneous Markov chain.

We also assume that $\mathcal{W}$ depends on the hidden Markov chain in the following way:

$$W_{j+1} := \mu(Y_j) + Z_{j+1}\Big(\underline{\xi}(Y_j)\Big), \quad j = 0, \dots, N-1, \tag{1.1}$$

where $Z(\underline{\xi}) := \{Z_{j+1}(\underline{\xi}), j = 0, ..., N-1\}$ is a sequence of random variables whose distribution depends on a vector of parameters $\underline{\xi} := (\xi_1, \dots, \xi_p)$ and the Markov chain $\mathcal{Y}$. At time t_j, the value of Y_j denotes the regime realization of the interval $[t_j, t_{j+1})$. The parameters $\mu \equiv \mu(Y_j)$ and $\underline{\xi} \equiv \xi(Y_j)$ are allowed to depend on the current regime, and they take values from a pre-determined set

$$\mu(k) := \mu_k \quad \text{and} \quad \underline{\xi}(k) := \underline{\xi}_k,$$

where $\underline{\xi}_k := (\xi_{k,1}, \dots, \xi_{k,p}) \in \mathbb{R}^p$, $k \in \mathbf{S}$, are given vectors of parameters.

1.1.2 Lévy Processes

A large class of models of the form (1.1) can be created by assuming that $Z_1, ..., Z_N$ correspond to increments of a Levy process. They became popular because of two main reasons:

(i) They form a rich class of statistical models.

(ii) They provide a tractable mathematical framework for pricing and hedging of financial options.

Below we only provide some basic facts about Lévy processes, and refer to Tankov (2003) for a comprehensive presentation of their properties and applications in finance.

Definition 1.1.1. Let $\mathcal{L} = \{L_t, t \geq 0\}$ be a stochastic process. Then it is a Lévy process if it satisfies the following properties:

1. $L_0 = 0$ almost surely.
2. For any $0 \leq t_1 < t_2 < \cdots < t_n < \infty, L_{t_2} - L_{t_1}, L_{t_3} - L_{t_2}, \ldots, L_{t_n} - L_{t_{n-1}}$ are independent.
3. For any $0 \leq s < t < \infty, L_t - L_s$ is equal in distribution to L_{t-s}.
4. For any $\epsilon > 0$ and $t \geq 0, \lim_{h \to 0} P(|L_{t+h} - L_t| > \epsilon) = 0$.

For a given Levy process $\mathcal{L} := \{L_t,\ t \geq 0\}$, we can define a hidden Markov process of the form (1.1) by taking $Z_{j+1} = L_{t_{j+1}} - L_{t_j}$, for a sequence $t_0 \leq \cdots \leq t_N$. Denote the characteristic function of $Z(\underline{\xi})$ by $\Psi(\cdot; \underline{\xi})$, which we assume to have an analytic representation. This assumption implies that conditional on Y_j, the characteristic function of the process W_{j+1} can be determined from the characteristic function of $Z_{j+1}(\underline{\xi})$. By the definition of characteristic functions,

$$\Psi(z; \underline{\xi}) = E(e^{izZ(\underline{\xi})}), z \in \mathbb{R}.$$

Lévy-Khintchine representation

By the Lévy-Khintchine representation, the characteristic function of $\mathcal{L}$ can be written in the form of

$$\Psi_{\mathcal{L}}(z, t) := E[e^{iz \cdot L_t}] = e^{t\psi_{\mathcal{L}}(z)}, \tag{1.2}$$

where the characteristic exponent $\psi_{\mathcal{L}}$ is given by

$$\psi_{\mathcal{L}}(z) := i\gamma z - \frac{\sigma^2}{2} z^2 + \int_{\mathcal{R}} (e^{izx} - 1 - izx1_{|x|\leq 1})\nu_{\mathcal{L}}(dx), \tag{1.3}$$

and $i := \sqrt{-1}$.

Equation (1.3) implies that the process is fully determined by two real parameters, γ and σ, and a measure $\nu_{\mathcal{L}}$, which is called the Lévy measure of the process. The measure $\nu_{\mathcal{L}}$ is defined by

$$\nu_{\mathcal{L}}(A) := E\Big[\#\{t \in [0,1] : \ \Delta L_t \neq 0, \Delta L_t \in A\}\Big],$$

where A is a measurable set and $\Delta L_t := L_t - \lim_{s \to t^-} L_s$. $\nu_{\mathcal{L}}$ can be interpreted as it assigns to every measurable set A, the expected number of jumps per unit time, whose sizes belong to A. Different selections of the Lévy measure lead to processes with different characteristics, and below we present three examples:

(i) $\nu_{\mathcal{L}} \equiv 0$. Then $\mathcal{L}$ is a Brownian motion with drift γ and diffusion coefficient σ.

(ii) $\bar{\nu}_{\mathcal{L}} := \int_{\mathcal{R}\setminus\{0\}} \nu_{\mathcal{L}}(x)dx < \infty$ and $d\nu_{\mathcal{L}}/\bar{\nu}_{\mathcal{L}}$ admits a density function g with respect to the Lebesgue measure. In this case, $\mathcal{L}$ can be represented as a sum of a linear trend, a Brownian motion with diffusion coefficient σ, and a compound Poisson process with jump intensity $\bar{\nu}_{\mathcal{L}}$ and jump-size density function g.

(iii) $\int_{\mathcal{R}\setminus\{0\}} \nu_{\mathcal{L}}(x)dx = \infty$ and $\sigma = 0$. In this case, $\mathcal{L}$ has infinite number of jumps over any finite time interval and is referred to as a process with infinite activities.

We use increments of a Lévy process to define the variables $Z_1, \dots, Z_N$ in (1.1), and since μ is a location parameter, to avoid redundancy, we assume that the parameter γ in (1.3) is zero. Next, we provide three different hidden Markov models of the form (1.1) corresponding to different selections of the measure $\nu_{\mathcal{L}}$.

Log-normal model. Hardy (2001) has proposed a regime-switching model where log-returns of the underlying asset follow a normal distribution with the mean and variance taking constant values in each regime. It is equivalent to our model where $\underline{\xi} = \sigma$ and $Z(\sigma) \sim N(0, \sigma^2 h)$, or $Z_1, \dots, Z_N$ in (1.1) are increments of a Brownian motion (BM) over intervals of length h, and its drift term is zero and the diffusion term is σ. Since in this case the characteristic exponent of $\mathcal{L}$ is of the form

$$\psi_{\mathcal{L}}(z) = -\frac{z^2\sigma^2}{2},$$

the log-characteristic function of $Z(\sigma)$ is given by

$$\psi_{BM}(z;\sigma) := -\frac{z^2\sigma^2}{2}h. \tag{1.4}$$

Merton (M) model. Merton (1976) proposed an extension of the log-normal model by adding jumps in the asset price. The jumps follow a Poisson process with a constant intensity λ and the logarithm of the jump size follows a normal distribution with mean μ_J and standard deviation σ_J upon a jump. In this case, the characteristic exponent of $\mathcal{L}$ is

$$\psi_{\mathcal{L}}(z) = -\frac{z^2\sigma^2}{2} + \lambda(e^{-\frac{z^2\sigma_J^2}{2}+i\mu_J z} - 1).$$

In principle, by taking $\underline{\xi} := (\sigma, \lambda, \mu_J, \sigma_J)$, we may allow any of the four parameters to change in each regime. Then, the log-characteristic function of $Z(\underline{\xi})$ will take the form

$$\psi_M(z;\underline{\xi}) := \Big[-\frac{z^2\sigma^2}{2} + \lambda(e^{-\frac{z^2\sigma_J^2}{2}+i\mu_J z} - 1)\Big]h.$$

The exponential moments of this model are finite, though the tails of the process are heavier than those of Gaussian.

Variance gamma (VG) model. This model, where $\mathcal{L}$ has a variance gamma process, can be represented as a subordinated Brownian motion, that is, $L_t = W_{h_t}$, where W is a Brownian motion with drift θ and diffusion σ, and the subordinator h_t is a gamma process with the variance parameter ν, $\nu > 0$ (see Madan et al., 1998, and, for an application to ratchet EIAs, Ballotta 2010). The characteristic exponent of the process $\mathcal{L}$ is of the form

$$\psi_{\mathcal{L}}(z) = -\frac{1}{\nu}\ln(1 - i\theta\nu z + \frac{1}{2}z^2\sigma^2\nu),$$

from which we can find

$$\psi_{VG}(z;\underline{\xi}) := -\frac{h}{\nu}\ln(1 - i\theta\nu z + \frac{1}{2}z^2\sigma^2\nu)$$

with $\underline{\xi} := (\sigma_0, \theta_0, \nu)$. Both the Lévy density and the probability density function of increments of the variance gamma process have exponential tails with decay rates depending on the three parameters θ, σ and ν.

Other possible selections of the Lévy process $\mathcal{L}$ can be:

(i) **Kou's model**: The jump-diffusion model proposed by Kou (2002), which is the same as Merton's model except that the logarithm of the jump size has a double exponential distribution.

(ii) **Other infinite-activity Lévy processes**: Normal inverse Gaussian process (for example, Barndorff-Nielses 1997, 1998) and Tempered stable process (for example, Carr et al., 2003).

In all these cases, the characteristic functions of the increments of the processes are tractable analytically.

1.2 Problems and Contributions

In this section, we outline the problems that we would like to address in each of the following chapters and the contributions that we have made to these problems.

Chapter 2 We study the estimation problem for regime switching Lévy processes, where we extend the estimation method proposed by Feuerverger and McDunnough (1981) and Feuerverger (1990). In particular, we apply the discrete empirical characteristic function (DECF) estimation method that is based on joint characteristic functions of two consecutive observations (bivariate) to regime switching models in Lévy processes. We use the maximum-likelihood estimation method (Kim (1994)) as a benchmark, where density functions are approximated by truncated sums of Fourier cosine series expansions. Then we compare these two estimation methods in a numerical study.

We also discuss some computational challenges within the proposed estimation procedure, such as selection of the dimension of the characteristic function and selection of the number and the locations of the points at which a characteristic function is evaluated. To solve the former problem, we propose a method to select the dimension by comparing some conditional distributions derived from the model. To solve the latter problem, we propose two methods of selection of points. One method is based on finite difference approximations with optimization, where the objective function is based on a measure of singularity of a covariance matrix. The other method is based on a quantization method proposed by Pagés et al. (2004).

Chapter 3 We study the goodness-of-fit testing problem for regime switching models, where we extend two existing goodness-of-fit tests. One is a visual test, which is based on the test proposed by Altman (2004). Instead of distribution functions used in the paper, we use empirical characteristic functions for our visual test. This test can examine the fit of an estimated model to observations by plotting the empirical characteristic function versus the true one to test deviations from the reference line, which is a 45 degree straight line through the origin. It can also test the dependence structure between observations by plotting the multivariate empirical characteristic function versus the true one. The other test is an extension of the method proposed by Koutrovelis and Kellermeier (1981) to non-i.i.d. observations, where the test statistic is proven to follow the same chi-square distribution as in the i.i.d. case considered by the authors.

Chapter 4 We apply the estimation and testing methods proposed in Chapters 2 and 3 to the same set of S&P 500 real data as used in Hartman and Groendyke (2013), where we consider two-regime switching Black-Scholes, Merton and variance gamma models as our candidate models. The estimates obtained by the proposed DECF estimation method are comparable with those obtained by the authors. Based on the visual test results, we do not reject the estimated two-regime switching Black-Scholes model, but we reject the other two estimated models.

Chapter 5 We study the Bermudan option pricing problem for regime switching Lévy processes. We propose a numerical pricing method called the Projected Value (PV) method, where in a dynamic programming setup we calculate conditional expectations by first representing the current value of the option using a series expansion and then by applying the characteristic function. To see the advantages of the proposed pricing method, we compare it with two well-known methods, the COS method developed by Fang and Oosterlee (2008) and the least-squares (LS) method proposed by Carrière (1996) and Longstaff and Schwartz (2001). We also apply the PV method to the problem of pricing ratchet equity-indexed annuities (EIAs), which currently are popular products in insurance markets.

Chapter 6 We devise a method of sampling paths from conditioned processes for a two-regime switching Black-Scholes model. We apply the algorithm to the problem of pricing path-dependent options, where we use an Asian call option for illustrations. We also compare the pricing results obtained by the proposed algorithm and a commonly used method, called the forward method. Then we apply the proposed algorithm to the problem of static hedging of path-dependent options, where we extend the static hedging method and relevant theoretical results proposed by Kolkiewicz (2016) to our model. Finally, we briefly discuss another application of the conditioned process sampling algorithm, which is filtering the path of volatility given the terminal value of the price process of the underlying asset.

Chapter 7 We discuss extensions of the existing work and possible topics for future research.

Chapter 2

Estimation of Regime Switching Lévy Processes

2.1 Introduction and Motivation

In this chapter, we discuss some estimation methods based on characteristic functions for the regime switching model defined in (1.1).

To estimate regime switching models, maximum likelihood estimation(MLE) using the expectation-maximization(EM) algorithm and Bayesian estimation using the Markov Chain Monte Carlo(MCMC) are two commonly used methods. The EM algorithm was first applied to regime switching models by Hamilton(1990) and then refined by Kim (1994). In a Bayesian framework, Harris (1997) uses the Gibbs sampler, while Hardy (2002) uses the Metropolis Hastings Algorithm. Hardy (2001, 2002) and Hartman and Groendyke (2013) apply the MLE and Bayesian estimation methods to estimate regime switching Black-Scholes models in insurance markets, and Janczura and Weron (2013) use the MLE to estimate regime switching models in electricity markets.

The above methods are based on distribution or density functions, which are usually unknown or not easily tractable for Lévy processes. Thus, estimation methods based on characteristic functions are more suitable for processes whose characteristic functions are easier to compute than their density functions. The estimation approaches that are based on empirical characteristic functions are known as the ECF (empirical characteristic function) methods, and they have been applied to numerous models, including ARMA models (Knight and Yu (2002)), switching regression models (Xu (2010)), normal mixture models (Xu and Knight (2011)) and Ornstein-Uhlenbeck-based stochastic volatility models (Taufer et al. (2011)). Kotchoni (2012) uses characteristic functions within the generalized method of moments with a continuum of moments conditions (CGMM) to estimate stable distributions and autoregressive variance gamma models. These CGMM methods were previously proposed by Carrasco and Florens (2000) and Carrasco et al. (2007). Other relevant references to the ECF methods are Yu (1998) and Jiang and Knight (2002). A more comprehensive literature review is provided in Yu (2004).

The ECF methods can be continuous or discrete. An example of the first approach is the integrated squared error (ISR) estimation method proposed by Heathcote (1977), which minimizes a weighted integral of a distance between empirical and marginal characteristic functions. There is also a discrete version of Heathcote's estimation method called the $k - L$ procedure (for example, Feuerverger and McDunnough (1981)). The $k - L$ method is based on empirical and true characteristic functions evaluated at k points. It can be seen as a generalized representation of the maximum likelihood procedure in the Fourier domain.

Estimation problems can be also divided into various categories according to the assumed form of dependency of the data. Among the existing literature, Heathcote (1977), Feuerverger and Mureika (1977) and Koutrouvelis (1980) consider the case of independent and identically distributed (i.i.d.) observations, while Feuerverger and McDunnough (1981), Feuerverger (1990), Knight and Satchell (1996,1997) and Carrasco et al. (2002) discuss the non-i.i.d. case. The estimation approaches for the i.i.d. case are usually based on marginal empirical characteristic functions; however, the marginal empirical characteristic function may not identify all the parameters in the non-i.i.d. case. Therefore, joint characteristic functions (CFs) and conditional characteristic functions are considered for estimating in the latter case. Estimating regime switching models in Lévy processes belongs to the non-i.i.d. case. In this chapter, we only consider the approaches based on joint characteristic functions rather than conditional ones. This is motivated by the fact that the methods based on conditional characteristic functions are not directly applicable because of the presence of unobservable, or latent, state variables in our models, and those based on joint characteristic functions allow us to estimate model parameters directly without filtering out the realizations of the latent variables. Though this may result in the loss of efficiency in the estimation procedure.

When using the ECF methods, we have to address several computational challenges, such as proper selections of the dimension of the characteristic function, the weight function and the points at which a characteristic function is evaluated. These problems are very important; for example, the number and the locations of the points at which a characteristic function is evaluated can affect the accuracy and efficiency of the estimation results and they may also result in a singular covariance matrix that is used in the estimation procedure. Indeed, this problem is still an open question, though quite a few studies have discussed it. Koutrouvelis (1980) only gives some basic ideas on how to choose the points. Feuerverger and McDunnough(1981) and Feuerverger (1990) only mention that the points should be properly selected. Later, Knight and Satchell (1997) propose a generalized method of moments (GMM) scheme to select the points. In addition, Schmidt (1982) advocates a similar approach where moment generating functions are used instead of characteristic functions. Yu (1998) discusses other computational problems that may occur during the implementation of the ECF estimation procedure, such as the singularity problem of the covariance matrix.

In this chapter, we consider using the discrete instead of the continuous ECF methods to estimate our models. By using the continuous ECF methods, we can match all the moments of the ECF and theoretical CF continuously, and hence more information from the observations is used. Also we do not need to be concerned with the selection of different grids as they are integrated out provided that the weight function is chosen appropriately. However, there are some potential problems with the continuous ECF methods; for example, the selection of the weight function as mentioned above. In theory, the optimal weight function is obtained as the inverse Fourier transform of the score function (see, for example, Feuerverger and McDunnough (1981)). Note that the score function depends on the density function, which is usually unknown for a regime switching diffusion process. Thus, we usually use some methods to approximate the weight function in the implementation, which can affect the efficiency of the continuous ECF methods. The two most common choices for the weight function are exponential and normal, but using these two functions does not result in an efficient estimator, though the asymptotic properties are still preserved. Although, we face several computational challenges when implementing the discrete ECF methods. In the following sections, we propose some methods to solve these problems.

The fact that some aspects of the ECF methods must be selected judiciously makes them more difficult to use in practice. In this chapter, we plan to propose and investigate more objective ways of estimating models described in terms of their characteristic functions. To estimate model (1.1), we utilize an estimation procedure that matches a joint empirical characteristic function with the theoretical one over a discrete set of points. For the problem of selection of the dimension of the characteristic function, we propose an approach that compares some conditional distributions derived from the model. For the problem of proper selection of the points at which a characteristic function is evaluated, we suggest a systematic way based on finite difference approximations. As an alternative approach, we consider a quantization method proposed by Pages et al. (2004). We compare our estimation results with those obtained by the EM algorithm, where in the latter we use cosine expansions to approximate density functions (the method is presented in Appendix 2.A).

This chapter is organized as follows. In Section 2.2, we present the discrete ECF estimation method and discuss some computational issues that need to be resolved. The problem of proper selection of the points at which a characteristic function is evaluated is discussed in Section 2.3. We illustrate the proposed methods through a simulation study in Section 2.4.

2.2 Discrete ECF Estimation Method

In this section, we discuss some estimation methods for regime switching models based on characteristic functions. Since marginal distributions of such processes are not sufficient to identify all the parameters of the models, we consider using joint characteristic functions. Methods of estimation based on characteristic functions evaluated over a set of finite points have been discussed, among others, by Feuerverger (1990), Knight and Satchell (1996, 1997), and Yu (1998).

Following Feuerverger and McDunnough (1981) and Feuerverger(1990), we present a discrete ECF estimation procedure for a non-i.i.d. stationary process. In the thesis, we refer to this as the DECF method. Below we first describe the DECF method. Then we discuss several computational challenges when implementing the method, followed by the description of the methods that we propose to address these issues.

Let $\mathcal{W} := \{W_j, j = 1, 2, ..., N\}$ be defined as in model (1.1). For convenience, we index the characteristic function of W_j given $Y_{j-1} = k$ by the parameter $\underline{\xi}_k$ as

$$\phi(z; \underline{\xi}_k) := E[e^{izW_j} | Y_{j-1} = k], z \in \mathbb{R}^+.$$

By conditioning, the marginal characteristic function $\Phi(z; \underline{\xi})$, of the process $\mathcal{W}$ can be represented as

$$\Phi(z; \underline{\xi}) := E[e^{izW_j}] = \sum_{k=1}^{K} \pi_k \phi(z; \underline{\xi}_k). \tag{2.1}$$

In the case where in the estimation procedure we use the m-dimensional marginal distribution, we define the following vectors: $\mathbf{z} := (z_1, ..., z_m) \in (\mathbb{R}^+)^m$ and $\mathbf{W_{j,m}} :=$

$(W_j, ..., W_{j+m-1}) \in \mathbb{R}^m$, where $j = 1, ..., N + 1 - m$. By the definition of our model, the characteristic function of $\mathbf{W_{j,m}}$ can be expressed as

$$\Phi_m(\mathbf{z}; \underline{\xi}) := E[e^{i\mathbf{z}\mathbf{W}'_{\mathbf{j,m}}}] = \sum_{k_1=1}^{K} \cdots \sum_{k_m=1}^{K} \pi_{k_1} p_{k_1,k_2} \cdots p_{k_{m-1},k_m} \phi(z_1; \underline{\xi}_{k_1}) \cdots \phi(z_m; \underline{\xi}_{k_m}), \quad (2.2)$$

where $'$ denotes the transpose of a vector. We refer to (2.2) as the m-dimensional characteristic function. When the parameters in (2.2) are estimated, we denote the corresponding characteristic function by $\hat{\Phi}_m(\mathbf{z}; \hat{\underline{\xi}})$.

In addition, we denote the m-dimensional ECF based on observations, $W_1, ..., W_N$, by

$$\bar{\Phi}_{m,N}(\mathbf{z}) := \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} e^{i\mathbf{z}\mathbf{W}'_{\mathbf{j,m}}}. \quad (2.3)$$

When $m = 1$, the one-dimensional ECF becomes:

$$\bar{\Phi}_N(\mathbf{z}) = \frac{1}{N} \sum_{j=1}^{N} e^{izW_j}, z \in \mathbb{R}^+. \quad (2.4)$$

Define ζ_m and $\zeta_{m,N}$ as $2q-$dimensional vectors,

$$(\zeta_m)' \equiv \Big(\zeta_m(\underline{\xi})\Big)' := \Big\{\Re_m(r_1), ..., \Re_m(r_q), \Im_m(r_1), ..., \Im_m(r_q)\Big\} \quad (2.5)$$

and

$$(\zeta_{m,N})' := \Big\{\Re_{m,N}(r_1), ..., \Re_{m,N}(r_q), \Im_{m,N}(r_1), ..., \Im_{m,N}(r_q)\Big\}, \quad (2.6)$$

where $\Re_m$ and $\Im_m$ are the real and imaginary parts of Φ_m, while $\Re_{m,N}$ and $\Im_{m,N}$ are the real and imaginary parts of $\bar{\Phi}_{m,N}$, and $r_1, ..., r_q$ are some selected different vectors from $(\mathbb{R}^+)^m$.

The basic idea of the DECF estimation method for the parameter $\underline{\xi}$ in model (1.1) is to minimize a weighted sum of the distance between $\bar{\Phi}_{m,N}$ and Φ_m. In particular, Feuerverger and McDunnough (1981) suggest to estimate the parameter $\underline{\xi}$ by minimizing the quadratic form of

$$(\zeta_{m,N} - \zeta_m)' \mathbb{D} (\zeta_{m,N} - \zeta_m), \quad (2.7)$$

where $\mathbb{D}$ is a weight function.

As suggested by the authors, we estimate the parameters in model (1.1) by minimizing the quadratic form of

$$L_{m,N} = (\zeta_{m,N} - \zeta_m)'(\Omega_m)^{-1}(\zeta_{m,N} - \zeta_m), \tag{2.8}$$

where $\Omega_m \equiv \Omega_m(\underline{\xi})$ is the covariance matrix of $\zeta_{m,N}$ and $(\Omega_m)^{-1}$ is its inverse.

Although numerous authors have considered the computational problems we need to face when implementing the DECF method, some of the problems are still open questions, for example, the selection of the q points at which a characteristic function is evaluated. In addition, some of the existing results are not applicable to our models directly. For example, though an explicit formula for the covariance matrix Ω_m for model (1.1) is available, it is not easy to be implemented in practice especially when observations are dependent and the dimension m is large. Therefore, in the following sections we would like to propose some methods to address these computational problems for our model, and the three main problems considered in this chapter are the evaluation of Ω_m, the selection of the dimension m and the selection of the q points.

Below we have several remarks for the estimating functions (2.7) and (2.8).

Remark[1] 2.2.1.

(i) Assume that $\hat{\underline{\xi}}$ is a consistent root of equation (2.7). Feuerverger and McDunnough (1981) show that if $\mathbb{D}$ in (2.7) is selected to be the inverse of the covariance matrix of $\zeta_{m,N}$, which is the estimating function (2.8), then $N \cdot var(\hat{\underline{\xi}})$ converges to the asymptotic Fisher information per observation.

(ii) The choice of points $r_1, ..., r_q$ affects the efficiency of the estimation procedure. Feuerverger and McDunnough (1981) also show that the estimation procedure can be as efficient as the maximum likelihood estimation is if the points $r_1, ..., r_q$ are chosen to be sufficiently fine and extended.

(iii) The behaviours of the processes $\Re_m$ and $\Im_m$ can provide some guidance for the selection of the points. Note that as $\mathbf{z} \to \mathbf{0}$, the processes $\Re_m$ and $\Im_m$ are uncorrelated, and $\Re_{m,N}$ and $\Im_{m,N}$ become their good estimates. Therefore, it is more reasonable to have the points to be close to zero. Based on our implementation, the estimation procedure works well even the points are not close to zero, say 50. In addition, the points $r_1, ..., r_q$ should be selected such that Ω_m is non-singular. The selection of these points are discussed further in Section 2.3.

[1]Thank Professor Tony Wirjanto for his comments that help me to formulate some of these remarks.

(iv) Because $\Re$ and $\Im$ are cosine and sine functions and they are bounded, and hence the estimating function (2.8) is also bounded. Therefore, the DECF estimation procedure is robust for a large class of models.

(v) The continuous form of the estimating function (2.7) is a weighted integral of the distance between the two characteristic functions $\bar{\Phi}_{m,N}$ and Φ_m. Although for this case we do not encounter the problem of selecting the q points, we have to decide the weight function, which is not a trivial problem especially for the case when the dimension m is large.

The covariance matrix for model (1.1) is presented in the lemma below.

Lemma 2.2.1. The covariance matrix Ω_m of the vector $\zeta_{m,N}$ defined in (2.6) can be represented as:

$$\Omega_m = \begin{pmatrix} \Omega_{m,RR} & \Omega_{m,RI} \\ \Omega_{m,IR} & \Omega_{m,II} \end{pmatrix}, \tag{2.9}$$

where the corresponding elements in the four sub-matrices are :

$$(\Omega_{m,RR})_{i,l} = E\Big[\Re_{m,N}(r_i)\Re_{m,N}(r_l)\Big] - \Re_m(r_i)\Re_m(r_l),$$

$$(\Omega_{m,RI})_{i,l} = E\Big[\Re_{m,N}(r_i)\Im_{m,N}(r_l)\Big] - \Re_m(r_i)\Im_m(r_l),$$

$$(\Omega_{m,IR})_{i,l} = E\Big[\Im_{m,N}(r_i)\Re_{m,N}(r_l)\Big] - \Im_m(r_i)\Re_m(r_l),$$

$$(\Omega_{m,II})_{i,l} = E\Big[\Im_{m,N}(r_i)\Im_{m,N}(r_l)\Big] - \Im_m(r_i)\Im_m(r_l),$$

and

$$\begin{aligned} E\Big[\Re_{m,N}(r_i)\Re_{m,N}(r_l)\Big] &= \frac{1}{2(N-m+1)}\Big(\Re_m(r_i+r_l)+\Re_m(r_i-r_l)\Big) \\ &+ \frac{1}{2(N-m+1)^2}\sum_{k=1}^{N-m}(N-m+1-k)\Big[\Re\psi_k(r_i,r_l) \\ &+ \Re\psi_k(r_i,-r_l)+\Re\psi_k(r_l,r_i)+\Re\psi_k(r_l,-r_i)\Big], \end{aligned}$$

$$
\begin{aligned}
E\Big[\Re_{m,N}(r_i)\Im_{m,N}(r_l)\Big] &= \frac{1}{2(N-m+1)}\Big(\Im_m(r_i+r_l)-\Im_m(r_i-r_l)\Big)\\
&+ \frac{1}{2(N-m+1)^2}\sum_{k=1}^{N-m}(N-m+1-k)\Big[\Im\psi_k(r_i,r_l)\\
&- \Im\psi_k(r_i,-r_l)+\Im\psi_k(r_l,r_i)-\Im\psi_k(-r_l,r_i)\Big],
\end{aligned}
$$

$$
\begin{aligned}
E\Big[\Im_{m,N}(r_i)\Re_{m,N}(r_l)\Big] &= \frac{1}{2(N-m+1)}\Big(\Im_m(r_i+r_l)-\Im_m(r_l-r_i)\Big)\\
&+ \frac{1}{2(N-m+1)^2}\sum_{k=1}^{N-m}(N-m+1-k)\Big[\Im\psi_k(r_l,r_i)\\
&- \Im\psi_k(r_l,-r_i)+\Im\psi_k(r_l,r_l)-\Im\psi_k(-r_i,r_l)\Big],
\end{aligned}
$$

$$
\begin{aligned}
E\Big[\Re_{m,N}(r_i)\Re_{m,N}(r_l)\Big] &= \frac{1}{2(N-m+1)}\Big(\Re_m(r_i+r_l)-\Re_m(r_i-r_l)\Big)\\
&+ \frac{1}{2(N-m+1)^2}\sum_{k=1}^{N-m}(N-m+1-k)\Big[-\Re\psi_k(r_i,r_l)\\
&+ \Re\psi_k(r_i,-r_l)-\Re\psi_k(r_l,r_i)+\Re\psi_k(r_l,-r_i)\Big],
\end{aligned}
$$

where

$$
\psi_k(r_i,r_l) = E\Big[e^{(ir_i\mathbf{W}'_{\mathbf{1,m}}+ir_l\mathbf{W}'_{\mathbf{k+1,m}})}\Big]. \tag{2.10}
$$

Proof: The proof is similar to that for stationary stochastic processes presented in Knight and Satchell (1995), and hence it is omitted. □

In the case when $m=2$, equation (2.10) can be written in an explicit form as below:

1. For $k=1$:

$$\begin{aligned}
\psi_1(r_i, r_l) &= E\Big[e^{(ir_i(W_1,W_2)' + ir_l(W_2,W_3)')}\Big] \\
&= E\Big[e^{(i(r_{i1},r_{i2})(W_1,W_2)' + i(r_{l1},r_{l2})(W_2,W_3)')}\Big] \\
&= E\Big[e^{(ir_{i1}W_1 + i(r_{i2}+r_{l1})W_2 + ir_{l2}W_3}\Big] \\
&= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} \\
&\cdot \quad \phi(r_{i1}; \underline{\xi}_{k_1})\phi(r_{i2}+r_{l1}; \underline{\xi}_{k_2})\phi(r_{l2}; \underline{\xi}_{k_3}).
\end{aligned} \tag{2.11}$$

2. For $k \geq 2$:

$$\begin{aligned}
\psi_k(r_i, r_l) &= E\Big[e^{(ir_i(W_1,W_2)' + ir_l(W_{k+1},W_{k+2})')}\Big] \\
&= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_1} p_{k_1,k_2} p^{(k-1)}_{k_2,k_3} p_{k_3,k_4} \\
&\cdot \quad \phi(r_{i1}; \underline{\xi}_{k_1})\phi(r_{i2}; \underline{\xi}_{k_2})\phi(r_{l1}; \underline{\xi}_{k_3})\phi(r_{l2}; \underline{\xi}_{k_4}).
\end{aligned} \tag{2.12}$$

As mentioned earlier, we face the following three main issues when implementing the DECF estimation method:

(i) Evaluation of the characteristic function ψ_k when $k \geq 2$. Formula (2.12) is not easy to use in practice for two reasons:

   - computation of the transition densities $p^{(k-1)}_{k_2,k_3}$ can be demanding, especially when k is large;
   - multiple summations result in extreme computational efforts, especially when m is large.

(ii) Selection of the dimension m of the vector $\mathbf{W_{j,m}}$.

(iii) Selection of q and the points $r_1, ...r_q$.

Below we discuss the above three problems in detail.

2.2.1 Approximation of the Characteristic Function ψ_k

In this section, we propose an approximation of the representation (2.12). Recall that if the stationary distribution π exists for a hidden Markov chain with the transition matrix $\mathcal{P} = [p_{kl}]$, then

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \pi_j, i, j \in 1, ..., K.$$

Thus, for any k[2]≥ 2, $p_{k_2,k_3}^{(k-1)}$ in (2.12) can be rewritten as

$$p_{k_2,k_3}^{(k-1)} = \pi_{k_3} + \epsilon_{k_3,k-1}, \tag{2.13}$$

where $\epsilon_{k_3,k-1}$ is the error term.

To approximate the right-hand side of formula (2.12), we can use the fact that for every $\epsilon > 0$, there exists a k^* such that for any $n \geq k^*$, we have $\bar{\epsilon}_{k_3,n} := |p_{k_2,k_3}^{(n)} - \pi_{k_3}| \leq \epsilon$. Therefore, we can approximate the transition probability $p_{k_2,k_3}^{(k-1)}$ in (2.12) with the following series

$$\hat{p}_{k_2,k_3}^{(n)} = \begin{cases} p_{k_2,k_3}^{(n)}, n \leq k^* - 1 \\ \pi_{k_3}, \qquad n \geq k^*. \end{cases} \tag{2.14}$$

Obviously, when $n \geq k^*$, the smaller the value of $\bar{\epsilon}_{k_3,n}$ is, the more accurate the above approximation of $p_{k_2,k_3}^{(k-1)}$ becomes. Indeed, the accuracy and efficiency of such approximation depend on the rate of convergence of the transition probability matrix $\mathcal{P}$ to its stationary distribution π. Next, we will describe this rate of convergence more formally.

Following Bremaud (1999), let U be the matrix of linearly independent eigenvectors in $\mathbb{R}^K$ and u_i be the i^{th} column of U, which is a left eigenvector of $\mathcal{P}$, and each u_i is normalized to having an L_2 norm equal to 1. Let D be the diagonal matrix of left eigenvalues of $\mathcal{P}$, namely $D = diag(d_1, ..., d_K)$, where each d_i is the corresponding eigenvalue of the eigenvector u_i. Let the eigenvalues be ordered such that $|d_1| > |d_2| \geq |d_3| \geq ... \geq |d_K|$. If $\mathcal{P}$ is a row stochastic matrix, its largest left eigenvalue is $d_1 = 1$.

By the eigen-value decomposition, we have

$$\mathcal{P} = UDU^{-1}.$$

[2]This k represents the number of steps $(k-1)$, which is different from the k in $\mathcal{P} = [p_{kl}]$, where k represents a state. Thus, hereafter in this section, we use $n = k - 1$ to represent the number of steps and $k_i, i = 1, ..., K$, represents the states.

Let $\mathcal{P}^{(n)}$ be the n-th power of the matrix $\mathcal{P}$ and $n \in \mathbb{Z}^+$. Then we have

$$\mathcal{P}^{(n)} = UD^nU^{-1}.$$

Consider a transition matrix with $K = 2$ as

$$\mathcal{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

In this case, we have $|d_1| = 1$ and $|d_2| = |1-\alpha-\beta|$. Note that the stationary distribution π for this $\mathcal{P}$ is

$$\pi = \Big[\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\Big].$$

Then, for any $n \in \mathbb{Z}^+$, $P^{(n)}$ becomes

$$\mathcal{P}^{(n)} = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & -\beta \end{pmatrix}.$$

and since $|1-\alpha-\beta| < 1$, we have

$$\mathcal{P}^{(\infty)} := \lim_{n\to\infty} P^{(n)} = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}.$$

Therefore,

$$\mathcal{P}^{(n)} = \mathcal{P}^{(\infty)} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & -\beta \end{pmatrix} = \mathcal{P}^{(\infty)} + \frac{(d_2)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & -\beta \end{pmatrix}$$

or equivalently,

$$\bar{\epsilon}_n := \mathcal{P}^{(n)} - \mathcal{P}^{(\infty)} = \frac{(d_2)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & -\beta \end{pmatrix},$$

where the error term $\bar{\epsilon}_n$ is a decreasing function of n since $|d_2| < 1$.

We have two results that can be inferred from the above derivations:

(i) Each row of $\mathcal{P}^{(\infty)}$ has the same form as the stationary distribution π.

(ii) The convergence rate of $\mathcal{P}^{(n)}$ to the stationary distribution π is exponential (power of n) and determined by the second largest eigenvalue.

Below we provides two simple examples to illustrate the above facts.

Example 2.2.1. When

$$\mathcal{P} = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix},$$

by the eigen-value decomposition, we get

$$U = \begin{pmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{pmatrix},$$

and

$$D = \begin{pmatrix} -0.2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, $|d_1| = 1$ and $|d_2| = 0.2$.

The stationary distribution π for this $\mathcal{P}$ is

$$\pi = [0.5 \;\; 0.5],$$

and for any $k \in \mathbb{Z}^+$,

$$\mathcal{P}^{(n)} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} + \frac{(-0.2)^n}{1.2} \begin{pmatrix} 0.6 & -0.6 \\ -0.6 & -0.6 \end{pmatrix},$$

where

$$\bar{\epsilon}_n = \frac{(-0.2)^n}{1.2} \begin{pmatrix} 0.6 & -0.6 \\ -0.6 & -0.6 \end{pmatrix}.$$

We find that $|\bar{\epsilon}_n| < 10^{-4}$ or $\mathcal{P}^{(n)}$ is equal to its stationary distribution up to the fourth decimal place when $n = 6$.

Example 2.2.2. When

$$\mathcal{P} = \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix},$$

by the eigen-value decomposition, we get

$$U = \begin{pmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{pmatrix},$$

and

$$D = \begin{pmatrix} -0.8 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, $|d_1| = 1$ and $|d_2| = 0.8$.

The stationary distribution π for this $\mathcal{P}$ is

$$\pi = [0.5 \ \ 0.5],$$

and for any $k \in \mathbb{Z}^+$,

$$\mathcal{P}^{(n)} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} + \frac{(-0.8)^n}{1.8} \begin{pmatrix} 0.9 & -0.9 \\ -0.9 & -0.9 \end{pmatrix},$$

where

$$\bar{\epsilon}_n = \frac{(-0.8)^n}{1.8} \begin{pmatrix} 0.9 & -0.9 \\ -0.9 & -0.9 \end{pmatrix}.$$

We find that $|\bar{\epsilon}_n| < 10^{-4}$ or $\mathcal{P}^{(n)}$ is equal to its stationary distribution up to the fourth decimal place when $n = 42$.

From these two examples, we can see that the $\mathcal{P}^{(n)}$ in Example 2.2.1 approaches π faster than that in Example 2.2.2, because the second largest left eigenvalue of Example 2.2.1 is smaller than that of Example 2.2.2.

The dimension of $\mathbf{W_{j,m}}$ used in the above results and examples is $m = 2$. However, the results can be extended to higher dimensions by the Perron-Frobenius Theorem (see Bremaud (1999)).

For convenience of reference, we summarize some of our findings in the following remark.

Remark 2.2.2.

(i) If $\mathcal{P}^{(n)}$ approaches π slowly, we can use (2.14) with a suitably larger value k^*.

(ii) If $\mathcal{P}^{(n)}$ approaches π fast, we can take $k^* = 1$, which makes the procedure faster and will not reduce the accuracy and efficiency a lot. In our implementation, we found that $k^* = 1$ was sufficient to produce reasonably accurate estimation results for our selections of the model parameters.

Now we use (2.14) to simplify (2.12).

Lemma 2.2.2. If $p_{k_2,k_3}^{(k-1)}$ in (2.12) is replaced by (2.14), then (2.12) becomes:

$$\bar{\psi}_k(r_i, r_l) = \begin{cases} \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3}^{(k-1)} p_{k_3,k_4} \\ \cdot\phi(r_{i1}; \underline{\xi}_{k_1})\phi(r_{i2}; \underline{\xi}_{k_2})\phi(r_{l1}; \underline{\xi}_{k_3})\phi(r_{l2}; \underline{\xi}_{k_4}),\ 2 \le k \le k^* \\ \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \pi_{k_1} p_{k_1,k_2} \phi(r_{i1}; \underline{\xi}_{k_1})\phi(r_{i2}; \underline{\xi}_{k_2}) \\ \cdot \sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_3} p_{k_3,k_4} \phi(r_{l1}; \underline{\xi}_{k_3})\phi(r_{l2}; \underline{\xi}_{k_4}), \quad k \ge k^* + 1, \end{cases} \tag{2.15}$$

and for any $\epsilon^* > 0$, there exists a k^* such that $\epsilon_{k,k^*} := |\psi_k(r_i, r_l) - \bar{\psi}_k(r_i, r_l)| \le \epsilon^*$.

Proof: See Appendix 2.B. □

The following lemma describes the special case when $p_{k_2,k_3}^{(k-1)}$ in (2.12) is replaced by (2.14) with $k^* = 1$.

Lemma 2.2.3. For any $k \ge 2$, if we assume that $\mathbf{W_{1,m}}$ and $\mathbf{W_{k+1,m}}$ are independent, then (2.12) is the same as (2.15) with $k^* = 1$.

Proof: See Appendix 2.C. □

If we substitute equations (2.11) and (2.15) into (2.10), then we get an approximation of the covariance matrix, which can be used in equation (2.8). In the rest of the chapter, we use (2.14) with $k^* = 1$.

2.2.2 Selection of the Dimension of the Vector $\mathbf{W_{j,m}}$

Another important implementational issue within the DECF estimation procedure is the selection of the dimension m. In general, larger m can increase the asymptotic efficiency of the DECF method, since more information about the model is included in the estimation procedure; at the same time, however, larger m makes an implementation more cumbersome as well as time-consuming.

Knight and Yu (2002) mention that a proper choice of m depends on the dimension of the minimal sufficient statistic of a model. Particularly, if the random variables of the observations are independent, then $m = 1$ is sufficient. If a model is a Markov process of order m^*, then $m = m^* + 1$ is sufficient, where $m^* = 1, 2, \ldots$. If a model is a non-Markov process with a sample size N, any $m < N$ is not sufficient. Intuitively, if a non-Markov process can be approximated well by a Markov process of order m^*, then $m = m^* + 1$ should be a good choice.

Our models are non-Markov processes, and $\{W_j, j = 1, ..., N\}$ are dependent variables because of the presence of the Markov chain $\mathcal{Y}$. This observation suggests that any choice of $m < N$ is not sufficient. Therefore, if m is small, estimation results may not be accurate; if m is large, the estimation procedure may take very long time and hence be inefficient. Thus, in our study, we only consider m up to 3.

Before we propose a procedure to select m, we need to describe in greater detail the estimating functions that we use in our model. Let $\underline{w} = \{w_1, ..., w_N\}$ be the observations drawn from the process $\mathcal{W} := \{W_j, j = 1, 2, ..., N\}$ defined in model (1.1). Since $\{W_j, j = 1, ..., N\}$ are not independent, the likelihood function based on the observations $\underline{w}$ is of the form

$$L(\underline{w}; \underline{\xi}) := f(w_1; \underline{\xi}) \prod_{j=2}^{N} f(w_j | \underline{w}_{j-1}; \underline{\xi}),$$

where $\underline{w}_j := (w_1, ..., w_j)$. Then, the corresponding score function is of the form

$$S(\underline{w}; \underline{\xi}) := \sum_{j=1}^{N} S_j(\underline{w}_j; \underline{\xi}), \tag{2.16}$$

where

$$S_1 = \frac{\partial}{\partial \underline{\xi}} \log f(w_1; \underline{\xi}) \quad \text{and} \quad S_j = \frac{\partial}{\partial \underline{\xi}} \log f(w_j | \underline{w}_{j-1}; \underline{\xi}).$$

For models whose density functions are unknown explicitly, we can consider an inference method based on a transform, like the characteristic function. In this case, we first need to specify the dimension m of a marginal distribution that we are going to use. Then an estimator $\hat{\underline{\xi}}$ can be defined as a solution to the equation

$$\sum_{j=1}^{N} G(\mathbf{W_{j,m}}; \hat{\underline{\xi}}) = 0, \tag{2.17}$$

with a suitably chosen function G. For independent observations, in which case $m = 1$ is sufficient, Feuerverger and McDunnough (1984) show that by properly selecting more points at which a characteristic function is evaluated, an estimating function (2.17) can be made close to (2.16) if $G(\mathbf{W_{j,m}}; \hat{\underline{\xi}})$ is close to the score function S_j in a properly selected L^2-space. For non-independent observations, which is our case, we are generally trying to approximate a function of N arguments using a function of only m $(m < N)$ arguments. To use this observation in practice, we have to first understand the dependence structure of our process.

There are several known facts about equations (2.17) and (2.18):

(i) If $\mathcal{W}$ is ergodic, then $\hat{\underline{\xi}}$ will converge to the true parameter $\bar{\underline{\xi}}$ that solves

$$E\Big[G(\mathbf{W_{1,m}}; \bar{\underline{\xi}})\Big] = 0. \tag{2.18}$$

(ii) If there is only one $\bar{\underline{\xi}}$ that solves (2.18), then the model is identifiable.

(iii) If the left-hand side (LHS) of (2.17) is asymptotically normally distributed, then so is $\hat{\underline{\xi}}$ (see, for example, Feuerverger and McDunnough (1984) and Feuerverger (1990)).

(iv) A practical way of determining the loss of efficiency of the method based on (2.17) when compared with the optimal estimating function (2.16) is still an open question.

In this section, we introduce a notion of closeness of two functions that depend on the same parameter. Define $f(x;\theta)$ and $g(x;\theta)$ as functions of x indexed by a parameter $\theta \in \mathbb{R}^d$, where $d \in \mathbb{Z}^+$. Let θ_0 be a fixed point. If for given x and $\epsilon > 0$, there exists a neighbourhood of θ_0 such that $|f(x;\theta) - g(x;\theta)| < \epsilon$, then we say that the two functions f and g are locally close, which we denote by $f(x;\theta) \overset{c.l}{\sim} g(x;\theta)$.

For the problem of proper selection of m, let us first look at an example of a Markov process of order $m^* = 1$. In this case, an estimation procedure based on the marginal distribution of the process $(m = 1)$ will retain the efficiency of the maximum-likelihood (ML) method only if

$$f(w_j; \underline{\xi}) \overset{c.l.}{\sim} f(w_j | w_{j-1}; \underline{\xi}), \quad j = 2, ..., N.$$

Typically, the above approximation will be poor, and hence we usually use transition densities $(m = 2)$ instead of the marginal distribution to estimate a Markov process.

If we choose $m = 2$, then it is possible to construct estimating functions based on marginal distributions of the process up to the dimension 2 that will lead to an estimating method with the same asymptotic efficiency as MLE. For this we can take

$$\bar{G}(w_j, w_{j-1}; \underline{\xi}) := (G_1(w_j, w_{j-1}; \underline{\xi}), G_2(w_j, w_{j-1}; \underline{\xi})),\ j = 1, ..., N,$$

with

$$G_1(w_j, w_{j-1}; \underline{\xi}) := \frac{\partial}{\partial \underline{\xi}} \log f(w_j; \underline{\xi}),\ j = 1, ..., N,$$

$$G_2(w_1, w_0; \underline{\xi}) := \frac{\partial}{\partial \underline{\xi}} \log f(w_1; \underline{\xi})$$

and

$$G_2(w_j, w_{j-1}; \underline{\xi}) := \frac{\partial}{\partial \underline{\xi}} \log f(w_j, w_{j-1}; \underline{\xi}),\ j = 2, ..., N.$$

Since

$$\log f(w_1|w_0; \underline{\xi}) = \log f(w_1; \underline{\xi})$$

and

$$\log f(w_j|w_{j-1}; \underline{\xi}) = \log f(w_j, w_{j-1}; \underline{\xi}) - \log f(w_{j-1}; \underline{\xi}),\ j = 2, ..., N,$$

it is easy to see that the estimating function

$$\sum_{j=1}^{N} \bar{G}(w_j, w_{j-1}; \underline{\xi}) = \mathbf{0}$$

is equivalent to the MLE, where $\mathbf{0}$ is a zero vector.

A similar analysis also holds if $\mathcal{W}$ is only weakly Markovian (i.e., the transition density is of the form $f(w_j|w_{j-1}, ..., w_{j-k}; \underline{\xi})$, for some $k \geq 1$). Then an estimation procedure based on (2.17) with $m < k + 1$ will have high efficiency if

$$f(w_j|w_{j-1}, ..., w_{j-m+1}; \underline{\xi}) \overset{c.l.}{\sim} f(w_j|w_{j-1}..., w_{j-k}; \underline{\xi}),\quad j = k + 1, ..., N. \tag{2.19}$$

If condition (2.19) holds, then, similar to the case when $\mathcal{W}$ is a Markov process, the score function corresponding to the conditional density function of the left-hand side of (2.19) can be recovered by properly defining estimating functions based on marginal distributions of the process up to the dimension m.

Below we present a procedure to select m for the model defined in (1.1), where we apply the above idea described for Markov models. Let $\underline{y} := \{y_1, ..., y_N\}$ be the realization of the hidden Markov process $\mathcal{Y} = \{Y_1, ..., Y_N\}$. In the following, we only consider two cases, $m = 1$ versus $m = 2$ and $m = 2$ versus $m = 3$, to decide which m (up to 3) is the best choice for our model.

Remark 2.2.3. From now on, we omit $\underline{\xi}$ for simplicity.

(i) $m = 1$ Versus $m = 2$.

By (2.19), we want to verify the condition

$$f(w_j) \overset{c.l.}{\sim} f(w_j|w_{j-1}), \quad j = 2, ..., N. \tag{2.20}$$

If (2.20) holds, then $m = 1$ is a sufficient choice; otherwise, we need to increase the value of m. For the LHS:

$$\begin{aligned} P(W_j = w_j) &= \sum_{y_j=1}^{K} P(W_j = w_j|Y_j = y_j)P(Y_j = y_j) \\ &= \sum_{y_j=1}^{K} P(W_j = w_j|Y_j = y_j)\pi_{y_j}. \end{aligned} \tag{2.21}$$

For the right-hand side (RHS):

$$\begin{aligned} P(W_j = w_j|W_{j-1} = w_{j-1}) &= \sum_{y_j=1}^{K} P(W_j = w_j|Y_j = y_j, W_{j-1} = w_{j-1})P(Y_j = y_j|W_{j-1} = w_{j-1}) \\ &= \sum_{y_j=1}^{K} P(W_j = w_j|Y_j = y_j) \sum_{y_{j-1}=1}^{K} p_{y_{j-1},y_j} P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}). \end{aligned}$$

Thus (2.20) holds if for any $j = 2, ..., N$, the following two conditions are satisfied:

- $P(W_j = w_j) = P(W_j = w_j|Y_j = y_j)$, which can be obtained from the fact that

  $$P(W_j = w_j) = P(W_j = w_j) \sum_{y_j=1}^{K} \pi_{y_j} = \sum_{y_j=1}^{K} P(W_j = w_j)\pi_{y_j} = \sum_{y_j=1}^{K} P(W_j = w_j|Y_j = y_j)\pi_{y_j}.$$

  This condition is not realistic, but could be approximately satisfied if the dependency on the regime in the hidden Markov process is weak.

- $\pi_{y_j} \overset{c.l.}{\sim} \sum_{y_{j-1}} p_{y_{j-1},y_j} P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1})$. This also could be approximately satisfied if the dependency on the regime is weak, in the sense that $P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}) \overset{c.l.}{\sim} \pi_{y_{j-1}}$.

These two conditions are very strong, and hence we typically use at least $m = 2$.

Since our models are non-Markovian, even $m = 2$ may not be sufficient. Therefore we may be interested in deciding whether there is any advantage in using $m = 3$, as opposed to $m = 2$.

(ii) $m = 2$ Versus $m = 3$.

By (2.19), we want to verify the condition

$$f(W_j = w_j|W_{j-1} = w_{j-1}) \overset{c.l.}{\sim} f(W_j = w_j|W_{j-1} = w_{j-1}, W_{j-2} = w_{j-2}), \quad j = 3, ..., N. \tag{2.22}$$

If (2.22) holds, $m = 2$ is a sufficient choice; otherwise, we may have to consider a larger value $m = 3$.

By Bayes' theorem, (2.22) can be equivalently represented as

$$\begin{aligned} &f(W_{j-2} = w_{j-2}, W_j = w_j|W_{j-1} = w_{j-1}) \overset{c.l.}{\sim} \\ &f(W_{j-2} = w_{j-2}|W_{j-1} = w_{j-1}) f(W_j = w_j|W_{j-1} = w_{j-1}), \quad j = 3, ..., N, \end{aligned} \tag{2.23}$$

which intuitively suggests that (2.22) will hold if W_{j-2} contributes less to our knowledge about W_j when W_{j-1} is known. In the extreme case when W_{j-1} is known, then W_{j-2} and W_j are independent, which is true if $\mathcal{W}$ is a Markov process of order $m^* = 1$. This also confirms our previous conclusions for $m = 2$ in the case of a Markov process of order $m^* = 1$. Conditions (2.22) and (2.23) also suggest that if we can filter the latent regime more accurately, then we are getting closer to a Markov process set-up, and hence (2.22) will be more likely true.

The following lemma provides a sufficient condition under which (2.23) holds.

Lemma 2.2.4. Condition (2.23) holds if for any $j = 3, ..., N$,

$$P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}) \overset{c.l.}{\sim} P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}, Y_{j-2} = y_{j-2}). \tag{2.24}$$

Proof: See Appendix 2.D.

Lemma 2.2.4 suggests that condition (2.23) holds if we only need the observation w_{j-1} to filter the state Y_{j-1} correctly or, equivalently, that conditional on the current observation the distribution of the hidden state does not change significantly if we add the information about the previous state.

Below we consider a two-regime switching model as examples to illustrate condition (2.24), where the state space of a Markov chain is $\mathbf{S} = \{1, 2\}$.

Through a simple algebra, we can represent the distributions in condition (2.24) in terms of model densities and parameters as presented in Lemma 2.2.5.

Lemma 2.2.5. Condition (2.24) holds if the following two conditions hold:

(i)

$$d_1 \overset{c.l.}{\sim} d_3 \overset{c.l.}{\sim} d_5, \tag{2.25}$$

where

$$d_1 := \frac{d_{W,1}p_{21}}{d_{W,1}p_{21} + d_{W,2}p_{12}},\ d_3 := \frac{d_{W,1}p_{11}}{d_{W,1}p_{11} + d_{W,2}p_{12}},\ d_5 := \frac{d_{W,1}p_{21}}{d_{W,1}p_{21} + d_{W,2}p_{22}},$$

with

$$d_{W,1} := P(W_{j-1} = w_{j-1}|Y_{j-1} = 1) \text{ and } \ d_{W,2} := P(W_{j-1} = w_{j-1}|Y_{j-1} = 2).$$

(ii)

$$d_2 \overset{c.l.}{\sim} d_4 \overset{c.l.}{\sim} d_6, \tag{2.26}$$

where

$$d_2 := \frac{d_{W,2}p_{12}}{d_{W,1}p_{21} + d_{W,2}p_{12}},\ d_4 := \frac{d_{W,2}p_{12}}{d_{W,1}p_{11} + d_{W,2}p_{12}},\ d_6 := \frac{d_{W,2}p_{22}}{d_{W,1}p_{21} + d_{W,2}p_{22}}.$$

Proof: See Appendix 2.E. $\square$

These two conditions correspond to taking y_{j-1} in (2.24) to be either 1 or 2 respectively. Let $\underline{\xi}_0$ be the true value of the parameter $\underline{\xi}$. We assume that the parameter $\underline{\xi}$ used for the distributions $d_{W,1}$ and $d_{W,2}$ is in the neighbourhood of $\underline{\xi}_0$. In the following lemma, we present a simple sufficient condition such that both conditions (2.25) and (2.26) hold. The proof is based on the following assumption. Define $g_1(w_{j-1};\underline{\xi})$ and $g_2(w_{j-1};\underline{\xi})$ as two functions of w_{j-1} and $T(\cdot)$ as a transformation of a function. We assume that the following condition holds at least for the neighbourhood of the parameter $\underline{\xi}_0$: if $g_1(w_{j-1};\underline{\xi}) \overset{c.l.}{\sim} g_2(w_{j-1};\underline{\xi})$, then $T(f(x;\underline{\xi})) \overset{c.l.}{\sim} T(g(x;\underline{\xi}))$, where $\underline{\xi}$ is in a neighbourhood of $\underline{\xi}_0$.

Lemma 2.2.6. Conditions (2.25) and (2.26) hold if the following condition holds:

$$g_{W,1} := \frac{d_{W,2}p_{12}}{d_{W,1}p_{11}} \overset{c.l.}{\sim} g_{W,2} := \frac{d_{W,2}p_{22}}{d_{W,1}p_{21}}, \tag{2.27}$$

or, more stringently, if the following condition holds:

$$\frac{p_{12}}{p_{11}} \overset{c.l.}{\sim} \frac{p_{22}}{p_{21}}. \tag{2.28}$$

Proof: See Appendix 2.F. □

Note that condition (2.28) is a sufficient condition for condition (2.27), and it does not depend on the distributions $d_{W,1}$ and $d_{W,2}$. Condition (2.28) implies that two rows of the transition matrix should be similar. Therefore, we can check conditions (2.25) and (2.26) by checking condition (2.28) instead. If condition (2.28) does not hold, then we need to check the less stringent condition (2.27).

Below we summarize the steps to decide m that is used in the DECF estimation procedure for a two-regime switching model based on the above results:

(S1–1) Estimate the model parameters by the DECF estimation method with $m = 2$ [3]. If $D_0 := \left|\frac{p_{12}}{p_{11}} - \frac{p_{22}}{p_{21}}\right| < \epsilon$, then $m = 2$ is a sufficient choice. Otherwise, we need to decide m by the following steps.

(S1–2) Estimate model parameters by the DECF estimation method with $m = 2$ and obtain the distributions $d_{W,j}$, $j = 1, 2$, by using the obtained parameters.

[3] In this section, we focus on the selection of the dimension m, but not on the selection of the best model. This step can be seen as a pre-analysis on the model parameters.

(S1–3) Obtain the normalized distributions of $g_{W,j}, j = 1, 2$, denoted by $g^*_{W,j}, j = 1, 2$, so that the distributions are comparable on the same scale. The normalizing constants can be obtained as $c_j = \sum_{\forall w_{j-1}} g_{W,j}$, then $g^*_{W,j} = \frac{g_{W,j}}{c_j}, j = 1, 2$.

(S1–4) If $D_1 := \max_{\forall w_{j-1}} \left| g^*_{W,1} - g^*_{W,2} \right| < \epsilon$, then $m = 2$ is a sufficient choice for the model; otherwise, it is more preferable to use $m = 3$.

Below we provide a numerical example to illustrate the above steps to decide the dimension m.

Example 2.2.3. In this example, we use $\epsilon = 0.01$, $w_{j-1} \in [-1, 1]$ and consider a two-regime switching Black-Scholes model, where the model parameters are $\mu_1 = 0.120$, $\sigma_1 = 0.114$, $\mu_2 = -0.141$, $\sigma_2 = 0.220$, $p_{12} = 0.045$, and $p_{21} = 0.149$. Since $D_0 := 5.66 > \epsilon$, condition (2.28) does not hold and we need to check condition (2.27) instead. Then we have $D_1 = 8.33 \cdot 10^{-16} < \epsilon$. Therefore, $m = 2$ is a sufficient choice for this set of model parameters.

We have repeated steps (S1–1)–(S1–4) for other values of parameters for the two-regime switching Black-Scholes model as well as for the two-regime switching Merton and variance gamma models. Below we list our main findings:

(i) Based on simulation studies, we find that the mean squared errors of the model parameter estimates with $m = 2$ on average are less than those with $m = 3$. This could be explained by the following arguments. If we consider MLE, an estimating function with $m = 2$ is based on the likelihood of d_1 (d_2), while for the case when $m = 3$ it is based on the likelihoods of d_3 and d_5 (d_4 and d_6). For most of the cases, we have $d_1 \overset{c.l.}{\sim} d_3 \overset{c.l.}{\sim} d_5$ ($d_2 \overset{c.l.}{\sim} d_4 \overset{c.l.}{\sim} d_6$). However, our estimation method is not MLE, so we may expect some cases such that the estimates with $m = 3$ on average perform better (smaller mean squared errors) than those with $m = 2$.

(ii) It is common to have the case such that condition (2.28) does not hold, while condition (2.27) holds. This could be explained by the fact that the ratio $\frac{d_{W,2}}{d_{W,1}}$ dominate the values of $g^*_{W,1}$ and $g^*_{W,2}$. In addition, if $f(x;\theta) \overset{c.l.}{\sim} g(x;\theta)$, then due to continuity, it is very likely that $f(x;\bar{\theta}) \overset{c.l.}{\sim} g(x;\bar{\theta})$, where $\bar{\theta}$ is a set of parameters in the neighbourhood of θ.

Based on our analysis, we use $m = 2$ in the DECF estimation procedure for our models and data sets.

2.3 Selection of Points

In this section, we discuss the problem of selection of points at which a characteristic function is evaluated in an estimation procedure. Selections of the number and the locations of the points are crucial to the accuracy and efficiency of the DECF estimation method. The existing literature offers only limited results to guide us through this problem. To address this issue, we propose two methods of selection of the points:

(i) Finite difference approximations with optimization (the FDAWO method).

(ii) Quantization method.

These two methods are presented in Sections 2.3.1 and 2.3.2 respectively.

2.3.1 The FDAWO Method

In this section, we propose a systematic way for selection of the points that are used in the DECF estimation procedure. We also briefly discuss the problem of selection of the number of points. Finally, we summarize the steps to obtain the points by our proposed method.

For a fixed q, $r_1, ..., r_q$ should be chosen such that Ω_m in (2.8) is non-singular. Instead of choosing these points randomly, we propose a first order finite difference approximation with optimization method, which we call the FDAWO method. The method is easily applicable since it involves finite derivatives of characteristic functions or empirical characteristic functions, which are in tractable forms for our models.

To simplify our presentation, we assume that the moment generating function for our model exists. It can be recovered from the characteristic function through the standard formula $M(z) = \Phi(-iz)$. Thus, to estimate the model parameters, we will use the empirical moment generating function, which we shall denote by $\bar{M}(z) := \bar{\Phi}_N(-iz)$, where $\bar{\Phi}_N(z)$ is the corresponding empirical characteristic function.

The FDAWO method, which can be used to determine the points that are used for the DECF estimation method, is based on the following two main steps:

(S2–1) Obtain a set of initial points by using a first order finite difference (FDA) method to approximate the derivatives of an empirical moment generating function.

(S2–2) Obtain a set of optimized points by solving an optimization problem with the initial points obtained in step (S2–1).

In step (S2–1), we use a systematic way based on observations to obtain a set of initial points. During the estimation procedure, we need the inverse of the covariance matrix Ω_m, which need to be non-singular. However, those initial points may generate a singular covariance matrix. To avoid this issue, we can get a set of optimized points by solving an optimization problem in step (S2–2). We explain these two steps in detail below when the dimension of a characteristic function is $m = 1$, and then we extend the method to other values of m.

- **Step (S2–1)**

  We explain how to use a finite difference method to obtain a set of initial q points, $r_1, ..., r_q$. Let $u(z) := \bar{M}(z)$, that is,

$$u(z) = \frac{1}{N}\sum_{j=1}^{N} e^{zW_j}, \tag{2.29}$$

  and let $r_0, ..., r_q$ be such that $0 = r_0 < r_1 < ... < r_q$ and $r_j \in \mathbb{R}^+, j = 1, ..., q$. Denote the i^{th} derivative of u at 0 by

$$u^{(i)} := \frac{\partial^{(i)} u(z)}{\partial z^{(i)}}|_{z=0}, \tag{2.30}$$

  since $u(z) = \bar{\Phi}_N(-iz)$, equation (2.30) becomes

$$u^{(i)} = \frac{1}{N}\sum_{j=1}^{N} (W_j)^i, \tag{2.31}$$

  which is the i^{th} empirical moment.

  Our objective is to get one-sided approximations to derivatives at a prescribed level of accuracy. Using the Taylor series expansion, we can obtain all the points $r_1, ..., r_q$ by using iteratively the following algorithm.

  For $j = 1, ..., q-1$, we have

$$u(r_j) = u(0) + \sum_{i=1}^{j+1} u^{(i)} \frac{r_j^i}{i!} + \underline{\epsilon}_{r_j}, \tag{2.32}$$

and

$$u(r_{j+1}) = u(0) + \sum_{i=1}^{j+1} u^{(i)} \frac{r_{j+1}^i}{i!} + \underline{\epsilon}_{r_{j+1}}, \tag{2.33}$$

where $\underline{\epsilon}_{r_j}$ and $\underline{\epsilon}_{r_{j+1}}$ are the error terms. Subtracting equation (2.32) from (2.33) gives

$$u(r_{j+1}) - u(r_j) = \sum_{i=1}^{j+1} u^{(i)} \frac{r_{j+1}^i - r_j^i}{i!} + \underline{\epsilon}_{r_j, r_{j+1}}, \tag{2.34}$$

where $\underline{\epsilon}_{r_j, r_{j+1}}$ is the error term. Then equation (2.34) can be rewritten as

$$\frac{(j+1)!(u(r_{j+1}) - u(r_j) - \sum_{i=1}^{j} u^{(i)} \frac{r_{j+1}^i - r_j^i}{i!})}{(r_{j+1}^{j+1} - r_j^{j+1})} = u^{(j+1)} + \underline{\epsilon}_{r_j, r_{j+1}}. \tag{2.35}$$

If the error term is small, then

$$u^{(j+1)} \approx \frac{(j+1)!(u(r_{j+1}) - u(r_j) - \sum_{i=1}^{j} u^{(i)} \frac{r_{j+1}^i - r_j^i}{i!})}{(r_{j+1}^{j+1} - r_j^{j+1})}. \tag{2.36}$$

To avoid obtaining the same points and/or too small intervals between the points, we find the next point by using previously obtained points as opposed to finding the best point for each derivative. Given all the previous obtained points $r_1, ..., r_j$ and a reasonable small value of $\underline{\epsilon}_{r_j, r_{j+1}}$, say 0.01, we can obtain the point r_{j+1} by solving the linear equation (2.35) with substituted equations (2.29) and (2.31). Therefore, we can obtain all the points $r_1, ..., r_q$ by the above algorithm.

We refer to the above method as the FDA (finite difference approximation) method and we call the selected points obtained by the FDA method the initial points.

Below we present an example to illustrate the FDA method when $j = 1$ and $j = 2$, i.e., to obtain the first two points r_1 and r_2.

For $j = 1$, we have

$$u(r_1) = u(r_0 + (r_1 - r_0)) = u(0) + u^{(1)} r_1 + \underline{\epsilon}_{r_1}, \tag{2.37}$$

where $\underline{\epsilon}_{r_1}$ is the error term. Then (2.37) can be rewritten as

$$\frac{u(r_1) - u(0)}{r_1} = u^{(1)} + \underline{\epsilon}_{r_1}. \tag{2.38}$$

Note that when the error term is small enough, then

$$u^{(1)} \approx \frac{u(r_1) - u(0)}{r_1}. \tag{2.39}$$

Given a small $\underline{\epsilon}_{r_1}$ and functions (2.29) and (2.31), we can obtain the only unknown variable r_1 by solving the linear equation (2.38).

For $j = 2$, we have

$$u(r_1) = u(0) + u^{(1)} r_1 + u^{(2)} \frac{r_1^2}{2} + \underline{\epsilon}_{r_1}, \tag{2.40}$$

and

$$u(r_2) = u(0) + u^{(1)} r_2 + u^{(2)} \frac{r_2^2}{2} + \underline{\epsilon}_{r_2}, \tag{2.41}$$

where $\underline{\epsilon}_{r_1}$ and $\underline{\epsilon}_{r_2}$ are the error terms.

Subtracting equation (2.40) from (2.41) gives

$$u(r_2) - u(r_1) = u^{(1)}(r_2 - r_1) + u^{(2)} \frac{(r_2^2 - r_1^2)}{2} + \underline{\epsilon}_{r_1,r_2}, \tag{2.42}$$

where $\underline{\epsilon}_{r_1,r_2}$ is the error term. (2.42) can be rewritten as

$$\frac{2(u(r_2) - u(r_1) - u^{(1)}(r_2 - r_1))}{(r_2^2 - r_1^2)} = u^{(2)} + \underline{\epsilon}_{r_1,r_2}. \tag{2.43}$$

If $\underline{\epsilon}_{r_1,r_2}$ is small enough, then

$$u^{(2)} \approx \frac{2(u(r_2) - u(r_1) - u^{(1)}(r_2 - r_1))}{(r_2^2 - r_1^2)}. \tag{2.44}$$

Given r_1 that obtained from equation (2.38) and a small $\underline{\epsilon}_{r_1,r_2}$, we can obtain the second point r_2 by solving the linear equation (2.43) with substituted equations (2.29) and (2.31).

- **Step (S2–2)**

  To ensure that the covariance matrix Ω_m used in the estimation is non-singular, we solve an optimization problem in this step, where we maximize or minimize a criterion that is a measure of singularity of the covariance matrix. Among all the possible criteria, the determinant and the eigenvalues are the most familiar ones. Knight and Satchell (1997) and Schmidt (1982) obtain the optimal points by minimizing the determinant of the asymptotic covariance matrix, which is due to the requirement that the points should be close to zero. However, Quandt and Ramsey (1978) argue that small points may result in a singular covariance matrix. To avoid the singularity problem, we propose to use the condition number instead of the determinant. The condition number is a measure of stability or sensitivity of a matrix to numerical operations and is also related to the singularity of a matrix, which is defined below.

Definition 2.3.1. Let A be an $n \times n$ matrix. Denote the norm of a matrix by $\|\cdot\|$. Then the condition number of the matrix A is defined as $\kappa(A) := \|A\| \cdot \|A^{-1}\|$. (See Golub and Van Loan (1996).)

Remark 2.3.1. Note that $\kappa(A) \geq 1$. If $\kappa(A) = 1$, A is non-singular; if $\kappa(A) = +\infty$, A is singular. If the condition number is closer to 1, the matrix is said to be well-conditioned; if the condition number is much greater than one, say greater than 30, the matrix is said to be ill-conditioned. See more details in Golub and Van Loan (1996) and Cheney and Kincaid (2012).

Let A be a square matrix. The following statements are equivalent:

(i) A is singular.

(ii) The columns of A are linear dependent.

(iii) The determinant of A is 0.

(iv) The condition number of A is $+\infty$.

To obtain the optimized points used for the estimation procedure, we propose to solve an optimization problem by minimizing the condition number of the covariance matrix Ω_m, where the initial points are obtained by the FDA method. There are four aspects that worth emphasizing for this optimization method:

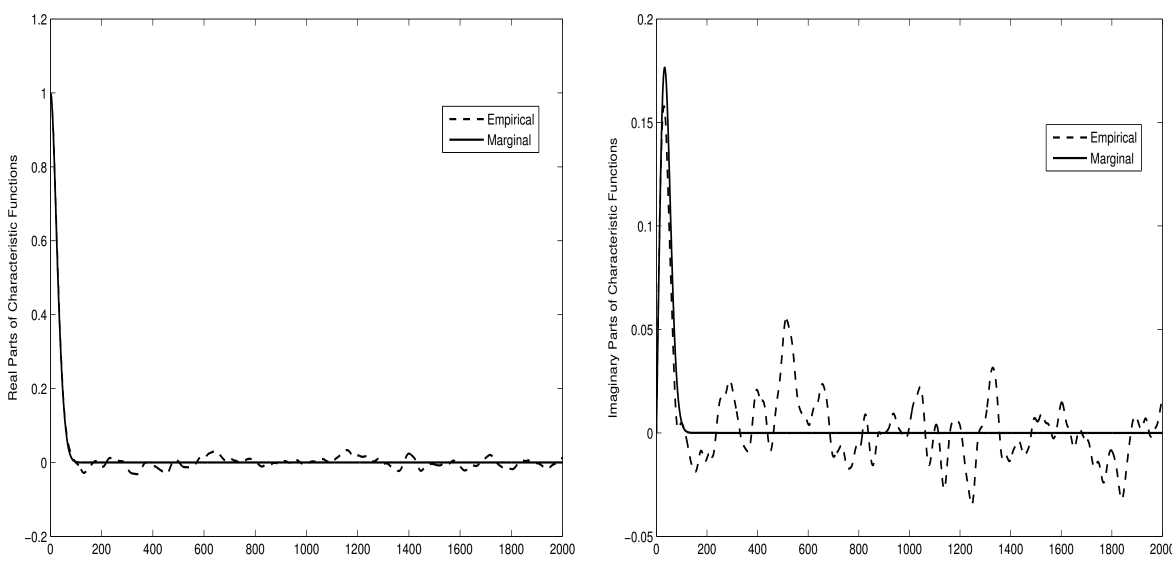
(i) Theoretically, we can also solve the optimization problem by minimizing the negative absolute value of the determinant (det) or any other criteria that are related to the covariance matrix. We use the condition number instead of the determinant as the criterion, because the optimization problem with the condition number has a boundary ($\min \kappa(A) = 1$) while the one with the determinant could be negative infinity or unbounded. Thus, using the condition number is more effective in our problem.

(ii) Minimization of the condition number is not a convex optimization problem (see Lu and Pong (2011)). However, our goal is to avoid singularity rather than finding the least singular matrix. Thus, we can obtain acceptable points at a local minimum of the condition number but not necessarily the globally best points.

(iii) Minimization of the condition number may result in points with large absolute values. However, when implementing the method, we can make the points bounded from above. The upper bound can be determined by imposing the condition that the modulus of the marginal characteristic function becomes less than a given number. For example, in the case presented in Figure 2.1, we can set the upper bound to be 100, where the real and imaginary parts of the marginal characteristic function become less than 0.

The following plot explains the reason for using an upper bound. Figure 2.1 shows the real (left panel) and imaginary (right panel) parts of the marginal and the empirical characteristic functions for an RSBS model, where we assume $m = 1$ and use the points $z = [0 : 1 : 2000]$. We can see that the marginal characteristic function is almost equal to zero when z is larger than 100. However, the empirical characteristic function is a trigonometric function, and it fluctuates around 0 permanently. Therefore, more information is included when z is not too large, while for large z, information contained in the empirical characteristic function is less reliable.

Figure 2.1: Characteristic Functions v.s Points $\mathbf{z}$



(iv) To avoid both the singularity problem and the extreme large values of points where the information is less reliable, we can bound the points in the optimization problem.

Below we extend the two steps (S2–1) and (S2–2) to the case when $m \geq 2$.

Definition 2.3.2. For $x \in \mathbb{R}$, the ceiling function of x is defined as the smallest integer not less than x and denoted by $\lceil x \rceil$.

When $m \geq 2$, we can obtain a set of initial points in three ways:

(i) First is a natural extension of the FDA method to higher dimensions. For a given q and m, choose q_* such that q_*^m is the smallest integer that greater or equal to q, that is, $q_*^m = \lceil q \rceil$. Then the m-dimensional q points are naturally to be $(r_1, ..., r_{q_*})^m$.

For example, if $q = 4$ and $m = 2$, then $q_* = 2$ and the $q = 4$ points used in the estimation are $(r_1, r_1), (r_1, r_2), (r_2, r_1)$ and (r_2, r_2), which form a rectangle.

(ii) Use $r_1, ...r_q$ to construct the q points such that they form a diagonal line through the origin.

For example, if $q = 4$ and $m = 2$, then the four points are $(r_1, r_1), ..., (r_4, r_4)$.

(iii) Select the coordinates of the points randomly from the points obtained by the FDA method. For example, if $q = 4$ and $m = 2$, and we have $2q = 8$ points $(r_1, ..., r_8)$ from the FDA method, then we can set the $q = 4$ points as $(r_1, r_2), (r_7, r_5), (r_4, r_6), (r_3, r_8)$ or any other permutations.

We refer to these three ways as Case 2 – (i), Case 2 – (ii) and Case 2 – (iii) respectively. To avoid the singularity problem of the covariance matrix, we select the optimized points by the same idea as presented in step (S2–2).

Another problem that we face when implementing the DECF method is the selection of the number of points, which is q. Theoretically, q should be large, since then we can retrieve more information about the model from its characteristic function, and the resulting estimator will be more efficient. On the other hand, larger number of points is likely to lead to a singular covariance matrix Ω_m. In our experience, q should at least satisfy $2q - p > 0$, and usually be no larger than three times of p, where p is the number of parameters that need to be estimated.

We summarize the steps for the FDAWO method to obtain the points used for the DECF estimation method as follows:

(S3–1) Choose the number q such that $q = \lceil \frac{p}{2} \rceil$ or $1 + \lceil \frac{p}{2} \rceil$, where p is the number of model parameters.

(S3–2) When $m = 1$, use the FDA method in equations (2.32)–(2.35) to obtain the initial points. When $m \geq 2$, use one of the three ways (i)–(iii) proposed in Case 2 to obtain the initial points.

(S3–3) Estimate the model parameters using the DECF estimation method with the initial points obtained in step (S3–2).

(S3–4) Set boundaries on the points when solving the optimization problem, and then obtain the optimized points used for the estimation procedure by minimizing the condition number of the covariance matrix, where the model parameters are estimated in step (S3–3).

2.3.2 The Quantization Method

In this section, we describe another way of selecting the points $r_1, ..., r_q$, which is based on the quantization method proposed by Pagès et al. (2004).

If the set of points $\mathbf{z}$ is chosen continuously, then we can estimate the model parameters by minimizing a distance between the joint characteristic function and the empirical one as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \Phi_m(\mathbf{z}; \underline{\xi}) - \bar{\Phi}_{m,N}(\mathbf{z}) \right|^2 g(\mathbf{z}) d\mathbf{z} \tag{2.45}$$

or

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \Phi_m(\mathbf{z}; \underline{\xi}) - \bar{\Phi}_{m,N}(\mathbf{z}) \right|^2 dG(\mathbf{z}), \tag{2.46}$$

where g and G are the weight functions. See Feuerverger (1990) and Yu (2004) for more details about continuous estimation approaches based on characteristic functions.

If the points are chosen to be finite and take at most q points, the approximation of the integral in (2.45) can be written as a Riemann sum with q terms as

$$\sum_{i=1}^{q} \left| \Phi_m(r_i; \underline{\xi}) - \bar{\Phi}_{m,N}(r_i) \right|^2 g_*(r_i), \tag{2.47}$$

where $r_1, ..., r_q \in \mathbb{R}^m$ and g_*[4] is a weight function. Thus, we may consider an alternative point selection method to select the q points, where the following difference is minimized:

$$\epsilon_{g,g_*} := \left| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \Phi_m(\mathbf{z}; \underline{\xi}) - \bar{\Phi}_{m,N}(\mathbf{z}) \right|^2 g(\mathbf{z}) d\mathbf{z} - \sum_{i=1}^{q} \left| \Phi_m(r_i; \underline{\xi}) - \bar{\Phi}_{m,N}(r_i) \right|^2 g_*(r_i) \right|. \tag{2.48}$$

Indeed, optimal quantization methods can solve this problem. The basic idea of quantization is to discretize the state space of a random vector or a stochastic process. It was originally used for discretizing emitted signal in 1950's, and since then it was developed and applied to information theory, signal processing and finance. For a complete review, we refer to Pagès et al. (2004) and Pagès and Printems (2008).

Optimal quantization of a $\mathbb{R}^d$ random vector $Z \in L^p(\Omega, \mathbb{P})$ addresses the problem of finding the best possible approximating random vector that takes at most N values such that the induced L^p-error is minimized. Following Pagès et al. (2004), let $X : \Omega \to \mathbb{R}^d$ be a random vector and $\Gamma = X(\Omega)$. If $\Gamma = \{x_1, ..., x_N\}$, then the L^p-quantization error is defined as

$$||Z - X||_p = E\Big[\min_{1 \le i \le N} |Z - x_i|^p \Big]. \tag{2.49}$$

Note that the expectation in equation (2.49) depends on the distribution of Z. Then the N-grid optimal quantizer of Z, $\{x_1, ..., x_N\}$, is obtained by minimizing equation (2.49).

In our case, we want to find a random vector that takes q values, $\Gamma = \{r_1, ..., r_q\}$, where $r_1, ..., r_q \in \mathbb{R}^m$, to replace the $\mathbb{R}^m$ random vector Z such that

$$E\Big[\min_{1 \le i \le q} |Z - r_i|^p \Big] \tag{2.50}$$

is minimized.

In our implementation, we use several specifications of the distribution of Z, which we explain below. The simplest one is the uniform distribution over an interval that is determined by the lower and upper boundaries used in the optimization problem. Once the points are determined, we can estimate the model parameters by minimizing (2.8) with the optimal quantizer Γ.

Since it is difficult to justify the use of a uniform distribution in (2.50) from the viewpoint of estimation errors, below we discus other choices of the distribution of Z. Feuerverger and McDunnough (1981, 1984) and Feuerverger (1990) demonstrate that with

[4] Note that $g_*(\cdot)$ is different from $g(\cdot)$, and it contains the approximation of $d\mathbf{z}$.

the DECF estimation method, $N \cdot var(\hat{\underline{\xi}})$ can be made arbitrarily close to the asymptotic Fisher information if the points are selected properly, where $\hat{\underline{\xi}}$ is a consistent root of equation (2.8). In addition, Feuerverger and McDunnough (1984) show that $N \cdot var(\hat{\underline{\xi}})$ can be written in the form of a weighted sum

$$\sum_{j=1}^{k+1} G_j^* \Big(\frac{\partial \log G_j^*}{\partial \underline{\xi}}\Big)^2, \tag{2.51}$$

where $G_j^* = G(r_j) - G(r_{j-1})$ and G is a CDF with $G(r_0) = 0$ and $G(r_{k+1}) = 1$. Thus (2.51) can be interpreted as a discrete approximation of the integral that appears in the definition of the Fisher information $I(\underline{\xi})$.

We would like to use this idea to select the q points. Specifically, Feuerverger and McDunnough (1981) show that a Fourier domain version of the likelihood function of MLE is in the form of

$$\int \Big(\Phi_m(\mathbf{z};\underline{\xi}) - \bar{\Phi}_{m,N}(\mathbf{z})\Big) g(\mathbf{z};\underline{\xi}) d\mathbf{z} = 0, \tag{2.52}$$

where $g(\mathbf{z};\underline{\xi})$ is a weight function, which is given as the inverse Fourier transform of the score function. The authors also mention that to solve (2.52), we can regard $g(\mathbf{z};\underline{\xi})$ as known. Then the estimating procedure is asymptotically equivalent to MLE if the weight function is appropriately selected. Note that if we take $g(\mathbf{z};\underline{\xi}) = g(\mathbf{z};\hat{\underline{\xi}})$, then the only component of integrand in (2.52) that depends on the unknown parameter $\underline{\xi}$ is $(\Phi_m(\mathbf{z};\underline{\xi}) - \bar{\Phi}_{m,N}(\mathbf{z}))$. The discrete approximation of equation (2.52) can be represented with a Riemann sum of q terms as

$$\sum_{i=1}^{q} \Big(\Phi_m(r_i;\underline{\xi}) - \bar{\Phi}_{m,N}(r_i)\Big) g(r_i;\underline{\xi}) = 0. \tag{2.53}$$

We would like to select the q points that lead to a fast convergence of the LHS of equation (2.53) to its limit, which is the LHS of equation (2.52). Since our estimation procedure is based on empirical characteristic functions, we can select the points $\mathbf{z} := \{r_1, ..., r_q\}$ by minimizing

$$\int \min_{1 \le i \le q} \Big| Z - r_i \Big|^p dG_N(Z), \tag{2.54}$$

where G_N is the empirical characteristic function of observations and hence is always positive. Similarly as in the FDAWO method, to avoid large values of the points, we can set boundaries when solving the optimization problem in the quantization method described above.

2.4 Simulation Study

In this section, we illustrate the DECF estimation method through a simulation study [5] for the model defined in (1.1). In Section 2.4.1, we consider a regime switching Black-Scholes model. In Section 2.4.2, we repeat our estimation procedures for a regime switching Merton model and a regime switching variance gamma model. In Section 2.4.3, we investigate the effects of different selections of points on estimation results.

2.4.1 Regime Switching Black-Scholes Model

In this section, the parameters we use for the simulation study are:

(i) $T = 2000$, the number of observations (values).

(ii) $N = 50$, the replication number for simulation.

(iii) Parameters in the FDAWO method:

- The error term used in the FDA method is 0.001.
- $q = 4$, the number of points used in the estimation.
- The lower and upper bounds on points are 0.1 and 100.

(iv) The model parameters for the RSBS model are the same as those used in Hartman Groendyke (2013). They are:

$$\mu_1 = 0.156,\ \sigma_1 = 0.110,\ \mu_2 = -0.096,\ \sigma_2 = 0.219,\ p_{12} = 0.045,\ p_{21} = 0.143.$$

The notations we use are:

- EM represents the modified EM algorithm for estimating regime switching models, which is presented in Appendix 2.A.
- Est1 represents the DECF estimation method with $m = 1$ when the points are selected by the FDAWO method.
- Est2i represents the DECF estimation method with $m = 2$ when the points are selected by the FDAWO method, and they form a rectangle as described in Case 2 – (i).

[5] All the numerical results in this thesis are obtained by using the software MATLAB.

- Est2ii represents the DECF estimation method with $m = 2$ when the points are selected by the FDAWO method, and they form a diagonal line through the origin as described in Case 2 – (ii).

- Est2iii represents the DECF estimation method with $m = 2$ when the points are selected by the FDAWO method, and they spread randomly as described in Case 2 – (iii).

The sets of points used in the following two methods are obtained by minimizing (2.54) when Z is uniformly distributed. (2.54) is calculated by a numerical integration method with N_L^m points, where N_L is a set of equally spaced points and m represents the m-fold Cartesian product. The initial points of $\mathbf{z}$ are randomly generated by the Sobol sequence generator. The number of the points and their lower and upper bounds are the same as those used in the FDAWO method.

- Est1Q represents the DECF estimation method with $m = 1$ and $N_L = 500$.

- Est2Q represents the DECF estimation method with $m = 2$ and $N_L^2 = 25^2$.

The resulting optimized points are:

Table 2.1: Selected Points for Estimation of RSBS

Method	r_1	r_2	r_3	r_4
Est1	21.2	49.5	78.7	100.0
Est2i	(25.1, 25.1)	(25.1, 75.1)	(75.1, 25.1)	(75.1, 75.1)
Est2ii	(22.0, 22.0)	(51.1, 51.1)	(78.7, 78.7)	(100.0, 100.0)
Est2iii	(35.2, 2.9)	(91.2, 42.3)	(41.3, 98.6)	(99.9, 100.0)
Est1Q	12.6	37.6	62.5	87.5
Est2Q	(25.1, 25.1)	(25.1, 75.1)	(75.1, 25.1)	(75.1, 75.1)

Remark 2.4.1. If the covariance matrix with the optimized points is still or close to be singular, we consider alternative ways to determine the initial points for the optimization problem:

(i) Drop the first or first two values of the initial points, then use the rest of the obtained points as the initials, for example, use $(r_2, r_3, ...)$.

(ii) Randomly select the needed number of points from the initial points as the new initial points. It is better to select the middle and large ones, and ensure that the distance between two numbers is big enough, for example, use $(r_2, r_4, r_5 ...)$.

Estimation results based on the selected points are presented in Table 2.2. The presented estimates are averages based on $N = 50$ repetitions, and the numbers in the parentheses are the corresponding standard deviations. In addition, Table 2.3 shows the mean squared errors of the estimates for different estimation methods.

We can see that the EM method shows the smallest bias, but for some parameters, their estimates based on the DECF method have much smaller standard deviations than the corresponding EM estimates. Thus, comparing the estimation methods using the mean squared error (MSE) [6] is more suitable. Based on the results presented in the tables, we have several findings as listed below:

(i) For most parameters when $m = 2$, their DECF estimates (Est2i, Est2ii, Est2iii and Est2Q) on average have much smaller MSEs than the corresponding EM estimates. Note that we use cosine series expansions to approximate density functions that are used in the EM algorithm, which may result in approximation and truncation errors. Since we focus on the DECF estimation method in this chapter, we only use the EM method as a benchmark. Therefore, the results at least confirm the accuracy of the DECF method.

(ii) Regardless of the way to select the points, the DECF estimates based on the joint characteristic function (Est2i, Est2ii, Est2iii and Est2Q) have much smaller MSEs than the corresponding ones based on the marginal characteristic function (Est1 and Est1Q). This agrees with our expectations because the joint characteristic function uses more information about the model than the marginal one does when estimating regime switching models.

(iii) When $m = 2$, the quantization method leads to estimates that perform no better than those based on the FDAWO method. This may be due to the fact that the FDAWO method depends on model parameters based on steps (S1-1)–(S1-5), while the quantization method used in this section does not, since here the distribution of $\mathbf{z}$ is assumed to be uniformly distributed in equation (2.54).

(iv) We have also found that estimates with randomly selected points (Case 2 – (iii)) have much smaller MSEs than those with the points that form either a rectangle (Case 2

[6] All the MSEs in this section are based on the simulation results.

– (i)) or a straight line (Case 2 – (ii)). This could be explained by the fact that the characteristic function with randomly selected points may contain more information of the model than the cases with points that are on a straight line or on the edges of a rectangle.

Table 2.2: True and Averaged Estimated Parameter Values of RSBS

	μ_1	σ_1	μ_2	σ_2	p_{12}	p_{21}
True	0.156	0.110	-0.096	0.219	0.045	0.143
EM	0.156	0.109	-0.092	0.220	0.047	0.153
	(0.015)	(0.003)	(0.047)	(0.010)	(0.012)	(0.034)
Est1	0.157	0.111	-0.468	0.178	0.049	0.294
	(0.015)	(0.010)	(0.499)	(0.048)	(0.025)	(0.257)
Est2i	0.153	0.112	-0.097	0.221	0.046	0.143
	(0.027)	(0.010)	(0.004)	(0.011)	(0.003)	(0.006)
Est2ii	0.157	0.110	-0.096	0.222	0.045	0.145
	(0.008)	(0.006)	(0.002)	(0.006)	(0.002)	(0.004)
Est2iii	0.157	0.110	-0.096	0.222	0.046	0.143
	(0.008)	(0.005)	(0.003)	(0.005)	(0.001)	(0.005)
Est1Q	0.155	0.109	-0.122	0.222	0.042	0.1205
	(0.016)	(0.006)	(0.123)	(0.019)	(0.026)	(0.075)
Est2Q	0.153	0.112	-0.097	0.221	0.046	0.143
	(0.027)	(0.010)	(0.004)	(0.011)	(0.003)	(0.006)

Table 2.3: Mean Squared Error of the Estimated Parameters of RSBS ($\times 10^{-4}$)

	μ_1	σ_1	μ_2	σ_2	p_{12}	p_{21}
EM	2	0	22	1	2	13
Est1	2	1	3871	40	7	884
Est2i	7	1	0	1	0	0
Est2ii	1	0	0	0	0	0
Est2iii	1	0	0	0	0	0
Est1Q	3	0	157	4	7	62
Est2Q	7	1	0	1	0	0

2.4.2 Other Regime Switching Models

In this section, we apply the DECF and EM estimation methods to a regime switching Merton model and a regime switching variance gamma model.

(i) The parameters of the RSM model are

$$\mu_1 = 0.237, \quad \mu_2 = 0.016, \quad \sigma_1 = 0.094, \quad \sigma_2 = 0.143, \quad \lambda_1 = 0.302, \quad \lambda_2 = 0.844,$$

and

$$\mu_{J1} = 0.521, \quad \sigma_{J1} = 0.450, \quad \mu_{J2} = -0.033, \quad \sigma_{J2} = 0.069, \quad p_{12} = 0.435, \quad p_{21} = 0.169.$$

They are obtained by estimating with the EM algorithm from the same S&P 500 data as used in Hartman and Groendyke (2013) [7].

(ii) The parameters of the RSVG model are obtained by matching prices of European put options with those obtained from the RSM model in (i). The options we use have one-year maturity and strikes that are equal to 80, 85, 90, 95, 100, 105, 110, 115, 120 and 125. The resulting parameters are:

$$\mu_1 = 0.151, \quad \mu_2 = 0.352, \quad \sigma_1 = 0.103, \quad \sigma_2 = 0.155, \quad \nu_1 = 0.0001, \quad \nu_2 = 0.024,$$

and

$$\theta_1 = -0.070, \quad \theta_2 = -0.303, \quad p_{12} = 0.417, \quad p_{21} = 0.179.$$

[7] Although, the measure has been changed from the real-world P to a risk-neutral Q, we only try to get a set of parameter values here for the estimation problem.

We only consider the cases EM, Est1, Est2iii, Est1Q and Est2Q for these two models. The selected points for the RSM and RSVG models are listed in Tables 2.4 and 2.5 respectively.

Table 2.4: Selected Points for Estimation of RSM

Method	r_1	r_2	r_3	r_4	r_5	r_6
Est1	0.5	63.2	70.7	82.3	95.7	100.0
Est2iii	(30.4, 12.3)	(35.2 ,60.2)	(86.4, 44.2)	(99.4, 99.9	(59.8, 99.5)	(98.3, 0.9)
Est1Q	8.4	25.1	41.7	58.4	75.0	91.7
Est2Q	(17.7, 17.7)	(24.2, 53.5)	(53.5, 24.3)	(75.7, 75.7)	(26.4, 84.9)	(84.9, 26.4)

Table 2.5: Selected Points for Estimation of RSVG

Method	r_1	r_2	r_3	r_4	r_5
Est1	18.1	40.9	63.0	86.0	100.0
Est2iii	(36.8, 0.6)	(0.1, 100.0)	(100.0, 51.2)	(100.0, 100.0)	(49.7, 79.7)
Est1Q	10.1	30.1	50.1	70.0	90.0
Est2Q	(22.1, 22.1)	(22.1, 78.0)	(77.9, 22.2)	(77.9, 77.9)	(50.0, 50.1)

The estimation results of these two models are presented in Tables 2.6 and 2.8. The estimates are averages based on $N = 50$ [8] repetitions, and the numbers in the parentheses are the corresponding standard deviations. Tables 2.7 and 2.9 show the mean squared errors of the estimates for each of the models. Based on the estimation results, we have the following findings:

(i) Similarly as reported in Section 2.4.1, the DECF estimates with $m = 2$ (Est2iii and Est2Q) on average have much smaller MSEs than those for the corresponding EM estimates. In addition, the DECF estimates based on the joint characteristic function (Est2iii and Est2Q) have much smaller MSEs than the corresponding ones based on the marginal characteristic function (Est1 and Est1Q).

(ii) Differently from those reported in Section 2.4.1, the case Est2Q leads to estimates that perform better than those of the case Est2iii. Indeed, we repeat the estimation procedure for models with different values of model parameters, and we find that

[8] We only use $N = 50$ because of the computational burden. It takes about 12 hours in total to obtain the estimation results for $N = 50$.

given m, the performances of the estimation method with points obtained by the FDAWO method (Case 2 – (iii)) and the quantization method are comparable and similar, though they depend on models and their values of parameters. Therefore, we recommend to use either of the two point selection methods to estimate those models with the DECF estimation method.

Table 2.6: True and Averaged Estimated Parameter Values of RSM

	μ_1	σ_1	λ_1	μ_{J1}	σ_{J1}	μ_2
True	0.237	0.094	0.302	0.521	0.450	0.016
EM	0.130	0.110	0.509	0.381	0.388	0.056
	(0.184)	(0.029)	(0.301)	(0.406)	(0.132)	(0.074)
Est1	0.318	0.114	0.293	0.471	0.430	0.435
	(0.078)	(0.008)	(0.019)	(0.027)	(0.010)	(0.069)
Est2iii	0.241	0.095	0.334	0.484	0.448	0.018
	(0.012)	(0.002)	(0.075)	(0.086)	(0.053)	(0.018)
Est1Q	0.289	0.089	0.953	0.238	0.421	0.024
	(0.164)	(0.047)	(0.611)	(0.544)	(0.240)	(0.301)
Est2Q	0.301	0.116	0.296	0.513	0.438	0.245
	(0.123)	(0.010)	(0.029)	(0.060)	(0.020)	(0.152)
	σ_2	λ_2	μ_{J2}	σ_{J2}	p_{12}	p_{21}
True	0.143	0.844	-0.033	0.069	0.435	0.169
EM	0.139	0.708	-0.130	0.079	0.395	0.177
	(0.022)	(0.546)	(0.106)	(0.030)	(0.169)	(0.143)
Est1	0.129	0.304	-0.039	0.066	0.418	0.156
	(0.011)	(0.027)	(0.050)	(0.003)	(0.021)	(0.011)
Est2iii	0.143	0.865	-0.035	0.069	0.435	0.168
	(0.003)	(0.086)	(0.011)	(0.011)	(0.014)	(0.006)
Est1Q	0.190	0.874	-0.300	0.101	0.437	0.187
	(0.154)	(0.680)	(0.266)	(0.094)	(0.174)	(0.169)
Est2Q	0.137	0.309	-0.017	0.068	0.422	0.159
	(0.011)	(0.044)	(0.061)	(0.006)	(0.028)	(0.015)

Table 2.7: Mean Squared Error of the Estimated Parameters of RSM ($\times 10^{-4}$)

	μ_1	σ_1	λ_1	μ_{J1}	σ_{J1}	μ_2	σ_2	λ_2	μ_{J2}	σ_{J2}	p_{12}	p_{21}
EM	455	11	1332	1843	213	384	26	4621	208	10	300	205
Est1	125	5	5	33	5	439	14	7	26	0	7	3
Est2iii	2	0	66	87	28	483	24	3234	1	1	2	0
Est1Q	297	22	7954	3753	583	1361	332	7887	1421	98	302	288
Est2Q	193	6	9	36	5	232	20	20	39	0	10	3

Table 2.8: True and Averaged Estimated Parameter Values of RSVG

	μ_1	σ_1	θ_1	ν_1	μ_2
True	0.151	0.103	-0.070	0.0001	0.352
EM	0.055	0.086	-0.066	0.007	0.322
	(0.305)	(0.038)	(0.269)	(0.013)	(0.171)
Est1	0.174	0.113	-0.071	0.0001	0.373
	(0.055)	(0.018)	(0.030)	(0.000)	(0.042)
Est2iii	0.122	0.104	-0.040	0.0001	0.370
	(0.070)	(0.002)	(0.070)	(0.000)	(0.022)
Est1Q	0.096	0.090	-0.060	0.008	0.299
	(0.270)	(0.039)	(0.261)	(0.018)	(0.141)
Est2Q	0.159	0.103	-0.066	0.0001	0.356
	(0.057)	(0.007)	(0.035)	(0.000)	(0.024)

	σ_2	θ_2	ν_2	p_{12}	p_{21}
True	0.155	-0.303	0.024	0.417	0.179
EM	0.157	-0.279	0.032	0.415	0.183
	(0.026)	(0.110)	(0.021)	(0.350)	(0.240)
Est1	0.155	-0.320	0.026	0.449	0.201
	(0.008)	(0.029)	(0.002)	(0.032)	(0.027)
Est2iii	0.156	-0.322	0.019	0.415	0.180
	(0.003)	(0.034)	(0.003)	(0.009)	(0.004)
Est1Q	0.159	-0.305	0.040	0.388	0.282
	(0.019)	(0.142)	(0.026)	(0.316)	(0.313)
Est2Q	0.154	-0.294	0.027	0.447	0.191
	(0.004)	(0.015)	(0.003)	(0.042)	(0.022)

Table 2.9: Mean Squared Error of the Estimated Parameters of RSVG ($\times 10^{-4}$)

	μ_1	σ_1	θ_1	ν_1	μ_2	σ_2	θ_2	ν_2	p_{12}	p_{21}
EM	1023	17	721	2	300	7	127	5	1225	575
Est1	35	4	9	0	22	1	11	0	20	12
Est2iii	57	0	57	0	8	0	15	0	1	0
Est1Q	761	17	682	4	226	4	201	10	1006	1084
Est2Q	34	0	12	0	6	0	3	0	26	6

2.4.3 Comparison of Different Methods of Point Selection

In this section, we compare the estimation results with different selections of points obtained through the quantization method. We consider two cases where the points are obtained by minimizing (2.54) when G_N is a uniform distribution and when G_N is the empirical characteristic function of observations. The initial points of $\mathbf{z}$ are generated by the Sobol sequence generator. In this section, we consider the same regime switching Black-Scholes model as used in Section 2.4.1.

Additional notations are:

- Est1unif represents the DECF estimation method with $m = 1$, $N_L^1 = 500$ and G_N is a uniform distribution.
- Est1ecf represents the DECF estimation method with $m = 1$, $N_L^1 = 500$ and G_N is the ECF of the simulated data.
- Est2unif represents the DECF estimation method with $m = 2$, $N_L^2 = 25^2$ and G_N is a uniform distribution.
- Est2ecf represents the DECF estimation method with $m = 2$, $N_L^2 = 25^2$ and G_N is the ECF of the simulated data.

The resulting points that are used for the estimation are presented in Table 2.10.

Table 2.10: Selected Points for Estimation of RSBS

Method	r_1	r_2	r_3	r_4
Est1unif	12.6	37.6	62.5	87.5
Est1ecf	20.8	55.5	31.3	87.8
Est2unif	(25.1, 25.1)	(25.1, 75.1)	(75.1, 25.1)	(75.1, 75.1)
Est2ecf	(15.2, 5.1)	(63.1, 66.7)	(15.9, 46.6)	(50.8, 7.3)

Table 2.11 shows the estimation results, where the presented estimates are averages based on $N = 50$ repetitions and the numbers in the parentheses are the corresponding standard deviations. In addition, Table 2.12 shows the mean squared errors for the estimates. Based on the results presented in these tables, we can conclude that

(i) For a given m, the estimates have much smaller MSEs on average when G_N is the empirical characteristic function than those when G_N is a uniform distribution. This agrees with our expectations, because we use more information of the model to select the points when G_N is the empirical characteristic function; while we do not use any model information to select the points when G_N is a uniform distribution.

(ii) The estimates based on the method Est2ecf have much smaller MSEs on average than those based on the method Est2unif. However, they perform no better than those of the case Est2iii in Section 2.4.1, and this result depends on models and their parameters as explained in Section 2.4.2. In addition, we left the study on the properties of the estimators based on different selections of the points for future research.

Table 2.11: True and Averaged Estimated Parameter Values of RSBS

	μ_1	σ_1	μ_2	σ_2	p_{12}	p_{21}
True	0.156	0.110	-0.096	0.219	0.045	0.143
EM	0.156	0.109	-0.092	0.220	0.047	0.153
	(0.015)	(0.003)	(0.047)	(0.010)	(0.012)	(0.034)
Est1unif	0.155	0.109	-0.122	0.222	0.042	0.1205
	(0.016)	(0.006)	(0.123)	(0.019)	(0.026)	(0.075)
Est1ecf	0.161	0.109	-0.091	0.224	0.047	0.156
	(0.016)	(0.005)	(0.020)	(0.023)	(0.012)	(0.083)
Est2unif	0.153	0.112	-0.097	0.221	0.046	0.143
	(0.027)	(0.010)	(0.004)	(0.011)	(0.003)	(0.006)
Est2ecf	0.168	0.094	-0.098	0.225	0.045	0.149
	(0.004)	(0.007)	(0.001)	(0.007)	(0.001)	(0.006)

Table 2.12: Mean Squared Error of the Estimated Parameters of RSBS ($\times 10^{-4}$)

	μ_1	σ_1	μ_2	σ_2	p_{12}	p_{21}
EM	2	0	22	1	2	13
Est1unif	3	0	157	4	7	62
Est1ecf	3	0	4	6	1	71
Est2unif	7	1	0	1	0	0
Est2ecf	2	3	0	1	0	1

2.5 Conclusions of Chapter 2

In this chapter, we present the DECF estimation method and discuss several computational challenges need to be resolved, including selections of the dimension of a characteristic function and the number and locations of the points at which a characteristic function is evaluated. Based on our analysis in Section 2.2.2, we use the dimension $m = 2$ for our models. To select the points, we propose two methods, one is based on finite difference approximations with optimization (the FDAWO method) and the other one is a quantization method proposed by Pages et. al (2004). We show that our proposed estimation and point selection methods work well through a simulation study, where we use the EM algorithm as a benchmark. Based on the numerical results in Section 2.4, we conclude that given the dimension m, the performances of the DECF estimation method with different selections of points depend on models and their parameters.

Chapter 3

Goodness-of-fit Testing for Regime Switching Models

3.1 Introduction and Motivation

In this chapter, we discuss some goodness-of-fit tests for the model defined in (1.1).

Besides estimation, model selection and goodness-of-fit (GOF) testing are another two important parts of statistical inference. Given a set of data, a model selection method can select a statistical model that fits the data best among all the candidate models, while a GOF test can individually describe how well a statistical model fits the data. Standard model selection methods, such as AIC, BIC and DIC, are widely used, and in the context of regime switching models have been applied, among others, by Hardy (2001) and Hartman and Groendyke (2013). These criteria independently measure the relative quality of a given set of models for the same observations, while parallel model selection methods, like the one used by Groendyke (2013), can simultaneously compare all the candidate models.

Regarding GOF testing, which is the focus of this chapter, we examine the fit of a selected model to a given set of observations. A well-known statistical test, Neyman-Pearson likelihood ratio (LR) test, usually compares a model with a nested submodel. This test has been applied to regime switching models in Hansen (1992), where the author shows that the LR statistic may have a conventional distribution under certain assumptions. However, there are several problems with the LR test used for regime switching models (see Date, Paresh and Mamon (2013)), for example, Hardy (2003) has stated that the LR test is not valid for the number of regimes in a regime switching model because of the asymptotics of the estimator for its test statistic. Another goodness-of-fit testing method is based on visual plots. For a stationary hidden Markov model, Altman (2004) has proposed such a test not only for univariate but also multivariate distributions. This method can be used to test the fit to the marginal distribution and the dependence structure of the underlying process. Since a visual test does not use any asymptotic distributions, it is a less rigorous method. All the tests mentioned above are not well suited for models defined by their characteristic functions, as they require knowledge of density functions. Thus, a goodness-of-fit test based on empirical characteristic functions is a more proper choice for testing models whose characteristic functions are easier to obtain than their density functions.

Koutrouvelis (1980) proposed a goodness-of-fit test for independent and identically distributed observations that is based on empirical characteristic functions when model parameters are given. This test has been extended by Koutrouvelis and Kellermeier (1981) to the case when model parameters need to be estimated. It is based on the same idea as the DECF method that we defined in Chapter 2, and it measures a distance between the empirical characteristic function and the true one at several pre-specified points. The test statistic has a chi-square distribution with a degree of freedom that depends on the number of points at which a characteristic function is evaluated and the number of parameters need to be estimated. The test is restricted to i.i.d. observations and requires proper selection of the number and locations of the points. Indeed, this test can be generalized by a class of minimum-distance methods based on empirical transforms proposed by Luong and Thompson (1987). Many papers have discussed continuous versions of this goodness-of-fit test, where the test statistics are weighted integrals of the deviations between the empirical characteristic function and the true one, like, Feuerverger and McDunnough (1984) and Jiménez-Gamero et. al (2009). Some other tests that are based on empirical characteristic functions have been applied to special distributions: Epps and Pulley (1983) propose a test for univariate normality, and Feuerverger and Mureika (1977) consider a test of univariate symmetry. These methods have been extended to multivariate tests for normality and symmetry by Ghosh and Ruymgaart (1992). However, these tests are only applicable to i.i.d. observations, so none of them is suitable for our models.

In this chapter, we present two extensions of the existing goodness-of-fit tests that are suitable for stationary hidden Markov models with known forms of characteristic functions. One is similar to the visual test proposed by Altman (2004), except that we use characteristic functions instead of distribution functions. By using this test, we can investigate the fit of the empirical characteristic function to the marginal one and also examine the dependence structure of observations. To ensure the convergence of the empirical characteristic function to the marginal one, some conditions must be satisfied. We apply the α-mixing result from Altman (2004) and the Lévy Continuity Theorem to show that the conditions hold for stationary hidden Markov models.

The second goodness-of-fit test that we propose is a natural extension of the test formulated by Koutrouvelis and Kellermeier (1981) to the case of non-i.i.d.. By using the α-mixing property of the process, we show that the test statistic has the same distribution as the one in Koutrouvelis and Kellermeier (1981). For this goodness-of-fit test, selection of points at which a characteristic function is evaluated is also as important as for the DECF estimation method discussed in Chapter 2. We demonstrate that we can use the methods proposed in Section 2.3 to select these points.

This chapter is organized as follows. We propose visual and statistical goodness-of-fit tests that are based on empirical characteristic functions for stationary hidden Markov models in Sections 3.2.2 and 3.2.3 respectively, followed by numerical illustrations in Section 3.3. Section 3.4 concludes.

3.2 Goodness-of-fit Testing

In practice we usually have several candidate models to consider. After estimating them, we want to select the best model among all the candidates and also test whether the estimated models fit the observations well. Indeed, we mainly focus on the goodness-of-fit testing for regime switching models or hidden Markov models in this chapter. A limited number of papers have discussed tests for such models, even less or none for models based on Lévy processes. To test such models, we have two things to consider. First, we have mentioned that their characteristic functions are easier than the density functions to compute, so tests based on characteristic functions are more reasonable. Second, visual tests are not rigorous though they intuitively give some information. Therefore, it is more desirable to have a formal test whose test statistic has a known distribution.

3.2.1 Method 1 – Visual Test

Altman (2004) has proposed a visual goodness-of-fit test for stationary hidden Markov models. The test visually shows the empirical against the estimated cumulative distribution functions in univariate and bivariate cases, and it examines deviations of the plots from the 45 degree (45°) line through the origin. The author not only assesses the fit of the theoretical distribution to the true one (the univariate case), but also analyses the dependence structure of observations (the bivariate case). In this chapter, we propose a similar test but based on empirical characteristic functions, where the model and notations are consistent with those used in Chapters 1 and 2.

Our test is based on a visual comparison of estimated univariate (bivariate) characteristic functions with empirical ones, $\hat{\Phi}_m(\cdot)$ versus $\bar{\Phi}_{m,N}(\cdot)$ at a range of points, where m =1 or 2, and it examines deviations of the plots from the 45° line through the origin, which we call the reference line. If the assumed model is correct or fits well to the observations, the plots are supposed to converge to the reference line.

Specifically, the x-axis of the plot is the real (or the imaginary) part of the empirical characteristic function evaluated at pre-specified points and the y-axis is the real (or the imaginary) part of the estimated characteristic function evaluated at the same points. Denote the real and imaginary parts of the estimated characteristic function by $\Re(\hat{\Phi}_m)$ and $\Im(\hat{\Phi}_m)$, and those of the empirical characteristic function by $\Re(\bar{\Phi}_{m,N})$ and $\Im(\bar{\Phi}_{m,N})$. Denote the number of points at which these functions are evaluated by q and the set of q points by $\{r_i, i = 1, ..., q\}$. Then we plot $\Re(\hat{\Phi}_m)$ against $\Re(\bar{\Phi}_{m,N})$ and $\Im(\hat{\Phi}_m)$ against $\Im(\bar{\Phi}_{m,N})$ respectively. If a model is estimated well, then the plots of $\Re(\hat{\Phi}_m)$ against $\Re(\bar{\Phi}_{m,N})$ and $\Im(\hat{\Phi}_m)$ against $\Im(\bar{\Phi}_{m,N})$ should lie on or show few deviations from the reference line; otherwise, the plots show obvious deviations from the reference line. For model selection purposes, we can compare all the candidate models by testing them on the same set of observations.

For the method to work, the following three assumptions are necessary.

(A2–1) $\mathcal{W}$ is strictly stationary. By definition, this requirement holds when the joint distribution of $(W_j, ..., W_{j+m})$ is identical for all j given a fixed m, where $m = 0, 1, ..., N - j$.

(A2–2) $\hat{\Phi}_m(\cdot)$ converges to $\Phi_m(\cdot)$ (in the sense of pointwise convergence). This requirement is satisfied, for example, when $\Phi_m(\cdot)$ is a continuous function of the model parameters and the parameters are estimated using a consistent method.

(A2–3) $\bar{\Phi}_{m,N}(\cdot)$ converges in probability to $\Phi_m(\cdot)$. To ensure that this is true, we can use the following Lemma.

Lemma 3.2.1. Let $\bar{\Phi}_{m,N}(\cdot)$ and $\Phi_m(\cdot)$ respectively be the empirical and the true characteristic functions of the random vector $\mathbf{W_{j,m}}$. Then $\bar{\Phi}_{m,N}(\cdot)$ converges to $\Phi_m(\cdot)$.

Proof: Let $\bar{F}_{m,N}(\cdot)$ and $F_m(\cdot)$ respectively be the empirical and the true cumulative distribution functions of the random vector $\mathbf{W_{j,m}}$. Proposition 1 in Altman (2004) shows the m-dimensional empirical cumulative distribution function converges to the true one, that is, $\bar{F}_{m,N}(\cdot)$ converges to $F_m(\cdot)$. By the Lévy Continuity Theorem, it is known that a sequence of random variables converges in distribution if and only if the corresponding sequence of characteristic functions converges to some function. Therefore, $\bar{\Phi}_{m,N}(\cdot)$ converges to $\Phi_m(\cdot)$. $\square$

Under conditions (A2–1)–(A2–3) above, our proposed visual test is valid for stationary hidden Markov models. Additionally, if $\bar{\Phi}_{m,N}$ converges in probability to Φ_m, then $\Re(\bar{\Phi}_{m,N})$ and $\Im(\bar{\Phi}_{m,N})$ converge in probability to $\Re(\hat{\Phi}_m)$ and $\Im(\hat{\Phi}_m)$ respectively. We conduct the visual test by plotting $\Re(\hat{\Phi}_m)$ against $\Re(\bar{\Phi}_{m,N})$ and $\Im(\hat{\Phi}_m)$ against $\Im(\bar{\Phi}_{m,N})$ to test the closeness between $\bar{\Phi}_{m,N}$ and $\hat{\Phi}_m$, where $m = 1$ and $m = 2$.

3.2.2 Method 2 – Statistical Test Based on the DECF Method

Koutrouvelis (1980) proposes a goodness-of-fit test using empirical characteristic functions where model parameters are given. Then Koutrouvelis and Kellermeier (1981) extend the test to the case where parameters need to be estimated first. Both tests assume i.i.d. observations. Since this assumption does not hold for hidden Markov models, we need to verify whether the test can be extended to our models.

First, let us assume $m = 1$, which implies $z \in \mathbb{R}^+$. Let $W_1, ..., W_N$ be random variables defined in (1.1) with a common characteristic function denoted by $\Phi(z)$. Recall equations (2.1) and (2.4) in Section 2.2.1, and define

$$\Phi_0(z, \underline{\xi}) := \Phi(z, \underline{\xi}) = \sum_{k=1}^{K} \pi_k \phi(z, \underline{\xi}_k), \tag{3.1}$$

$$\Phi_N(z) := \bar{\Phi}_N(z) = \frac{1}{N} \sum_{j=1}^{N} e^{izW_j}. \tag{3.2}$$

Define two $2q-$dimensional vectors:

$$\zeta_0^{'} \equiv \zeta_0^{'}(\underline{\xi}) := \Big\{ \Re_0(r_1), ..., \Re_0(r_q), \Im_0(r_1), ..., \Im_0(r_q) \Big\} \tag{3.3}$$

and

$$\zeta_N^{'} := \Big\{ \Re_N(r_1), ..., \Re_N(r_q), \Im_N(r_1), ..., \Im_N(r_q) \Big\}, \tag{3.4}$$

where $r_1, ...r_q$ are different points from $\mathbb{R}^+$ and $\Re_l$ and $\Im_l$ are the real and imaginary parts of Φ_l with $l = 0, N$. Let $\Omega_0 \equiv \Omega_0(\underline{\xi})$ be the $2q \times 2q$ covariance matrix as shown in (2.9) under the null hypothesis H_0. The arguments $r_1, ..., r_q$ can be chosen by the proposed point selection methods in Section 2.3 such that Ω_0 is non-singular. Suppose that we want to test that the estimated model is the true model. Then the null hypothesis of our test is:

$$H_0 : \Phi(z) = \Phi_0(z, \underline{\xi}), \tag{3.5}$$

where $\underline{\xi} := (\xi_1, \dots, \xi_p) \in \mathbb{R}^p$ is the set of parameters of our testing model, and $\underline{\xi}$ is an unknown parameter that needs to be estimated. Define $\hat{\underline{\xi}}_N$ as an estimate of this parameter obtained by minimizing equation (2.8) with $m = 1$. The test statistic under H_0 has the following quadratic form of:

$$Q_N^0 := (\zeta_N - \zeta_0)' \Omega_0^{-1} (\zeta_N - \zeta_0). \tag{3.6}$$

Before presenting some theoretical results for the proposed test, we need basic facts about α-mixing processes and the Central Limit Theorem for such processes.

Definition 3.2.1. Let $\mathcal{V} := \{V_j, j \in \mathbb{Z}^+\}$ be a stationary sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_a^b := \sigma\{V_j, a \le j \le b\}$ be the σ-algebra generated by the random variables $\{V_a, ..., V_b\}$. We say that $\mathcal{V}$ is α-mixing (or strong mixing) if

$$\alpha_l := \sup_{A \in \mathcal{F}_1^s, B \in \mathcal{F}_{s+l}^\infty} \Big| P(AB) - P(A)P(B) \Big| \to 0 \quad \text{as} \quad l \to \infty, \tag{3.7}$$

and $\{\alpha_l\}$ are called the mixing coefficients.

Bradley (1985) states the following Central Limit Theorem (CLT) for an α-mixing process.

Theorem 3.2.1. Suppose $\mathcal{X} = \{X_j, j \in \mathbb{Z}^+\}$ is strictly stationary, $E(X_1) = 0$, $\{\alpha_l\}$ are the mixing (α-mixing) coefficients of the centered process $\mathcal{X}$, and either of the following two conditions holds:

(i) for some $\delta > 0$, $E\Big(|X_1|^{2+\delta}\Big) < \infty$ and $\sum_{l=1}^{\infty} \alpha_l^{\frac{\delta}{2+\delta}} < \infty$,

(ii) for some $C < \infty$, $|X_1| < C$ a.s. and $\sum_{l=1}^{\infty} \alpha_l < \infty$.

If $\sigma_*^2 = E(X_1^2) + 2\sum_{j=2}^{\infty} E(X_1 X_j)$ and $0 < \sigma_*^2 < \infty$, then $\frac{\sum_{j=1}^{N} X_j}{\sqrt{N}\sigma_*} \to N(0, 1)$ in distribution as $N \to \infty$.

Let $\mathcal{V} := \{V_j, j \in \mathbb{Z}^+\}$ be a stationary hidden Markov process and $\mathcal{X}$ be the centered process $\mathcal{V}$, that is, $X_j = V_j - E(V_j), j \in \mathbb{Z}^+$. Then $\mathcal{X}$ is a stationary hidden Markov model, and hence we can apply Theorem 3.2.1 to the model defined in (1.1). Theorem 1 in Mackay (2002) states that the mixing coefficients for stationary hidden Markov models satisfy $\alpha_l = O(l^{-\nu})$ for some $\nu > 2q + 1$, where $q \in \mathbb{Z}^+$. Thus a stationary hidden Markov model is a α-mixing process by the definition of α-mixing.

Then we have the following theorem for the asymptotic distribution of the test statistic defined in equation (3.6).

Theorem 3.2.2. Assume that observations $\mathcal{W} = \{W_j, j \in \mathbb{Z}^+\}$ are from stationary hidden Markov models with finite centered moments and $\underline{\xi}_0$ is the true parameter under H_0, and $r_1, ..., r_q$ are positive real values such that $\Omega_0(\underline{\xi}_0)$ is non-singular. Then, under H_0,

$$Q_N := \Big(\zeta_N - \zeta_0(\hat{\underline{\xi}}_N)\Big)' \Omega_0^{-1}(\underline{\xi}_0)\Big(\zeta_N - \zeta_0(\hat{\underline{\xi}}_N)\Big) \tag{3.8}$$

has an asymptotic χ^2 distribution with $2q - p$ $(p < 2q)$ degrees of freedom as the sample size $N \to +\infty$.

Proof: See Appendix 3.A. □

Remark 3.2.1.

(i) The above result can be naturally extended to higher dimensions with $m \geq 2$ by using a multivariate CLT for α-mixing processes (see Tone (2010)). However, we find that an improper selection of the q points may result in an inaccurate distribution of the test statistic. We left the investigation of the implementation problems for this test in higher dimensions with $m \geq 2$ for future research.

(ii) When $m = 1$, an improper selection of the number q and the locations of the q points may result in a singular covariance matrix Ω_0, in which case the distribution of the test statistic could be different from the one described in Theorem 3.2.2. Thus, we propose to use the FDAWO or the quantization method introduced in Section 2.3 to select these points.

(iii) When the model parameters are given $(p = 0)$, the test statistic based on the quadratic form (3.8) is χ^2_{2q} distributed. Although the model parameters are usually unknown in practice, we can verify the accuracy of this test through a simulation study given the model parameters. The q points can also be obtained by the FDAWO or the quantization method except that now we use the known model parameters.

(iv) When the model parameters are unknown, they need to be estimated by the proposed DECF estimation method. Although we need to face the problem of selection of the grid points, we only consider the test based on the DECF estimator instead of the continuous ECF estimator in this chapter for consistency considerations.

(v) As suggested by my committee member, Professor Tony Wirjanto, in general, the result in Theorem 3.2.2 usually does not hold for weakly dependent observations with exception of mixing processes as we proved. In addition, when the model parameters are unknown, using the parametric bootstrap method (see Leucht (2012)) to obtain critical values and approximate p-values is an option. However, bootstrap samples may not be mixing even though the original process satisfies some mixing conditions, and hence the bootstrap method may not be appropriate for our models.

3.3 Simulation Study

For the proposed two goodness-of-fit tests, we use a set of simulated data where we estimate the parameters by the DECF method, and we want to test the performance of the two goodness-of-fit methods.

3.3.1 Method 1 – Visual Test

For the visual test, we use the same $q = 21$ points as those used in Altman (2004).

(i) For the univariate visual test ($m = 1$), $r_i = i - 1$ where $i = 1, ..., 21$.

(ii) For the bivariate visual test ($m = 2$), $r_i = (i - 1, i - 1)$ where $i = 1, ..., 21$.

To illustrate the proposed visual test, we only simulate one set of observations using the same regime switching Black-Scholes model as the one used in Section 2.4.1, and we denote the generated set of data by Data 1. Then we estimate the model parameters using Data 1 and the DECF method, where the points used in the estimation procedure are selected by the FDAWO method with Case 2 – (iii), and we denote this case by DECF2iii. The estimation results are presented in Table 3.1.

Table 3.1: Estimated Parameter Values

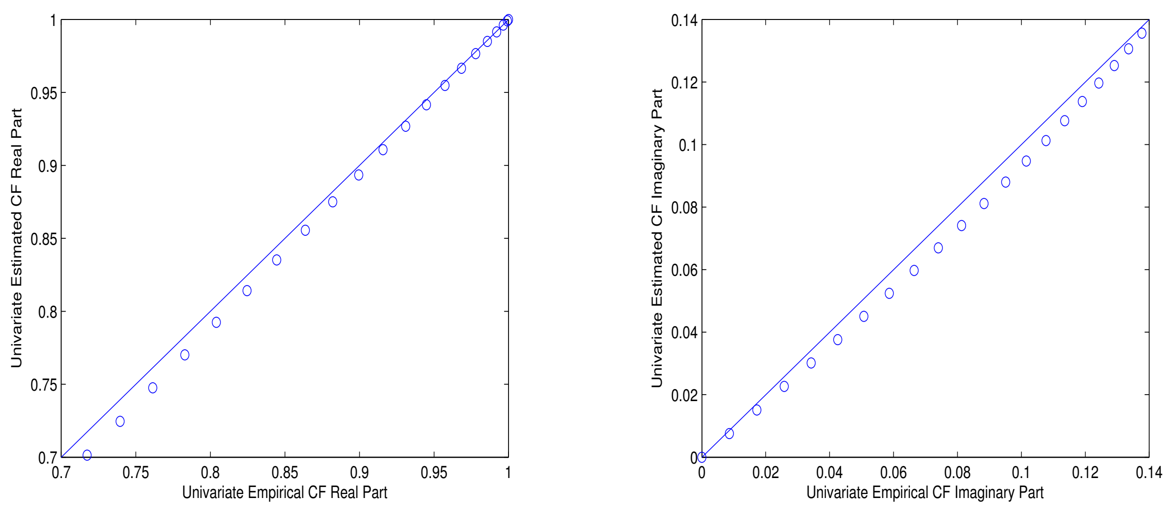
	μ_1	σ_1	μ_2	σ_2	p_{12}	p_{21}
True	0.156	0.110	-0.096	0.219	0.045	0.143
DECF2iii	0.153	0.112	-0.099	0.218	0.047	0.143

We now apply the visual test to examine the closeness of the estimated model and the true one by plotting $\Re(\bar{\Phi}_{m,N})$ versus $\Re(\hat{\bar{\Phi}}_m)$ and $\Im(\bar{\Phi}_{m,N})$ versus $\Im(\hat{\bar{\Phi}}_m)$ at the $q = 21$ points presented in (i) and (ii) above. We want to test deviations of the plots from the reference line (the 45 degree line through the origin), and we consider two cases: the univariate test when $m = 1$ and the bivariate test when $m = 2$.

Figures 3.1-3.2 show that all the plots mimic the reference lines well, which is consistent with the way data were generated. Therefore, our estimated model fit the true one well based on these figures.

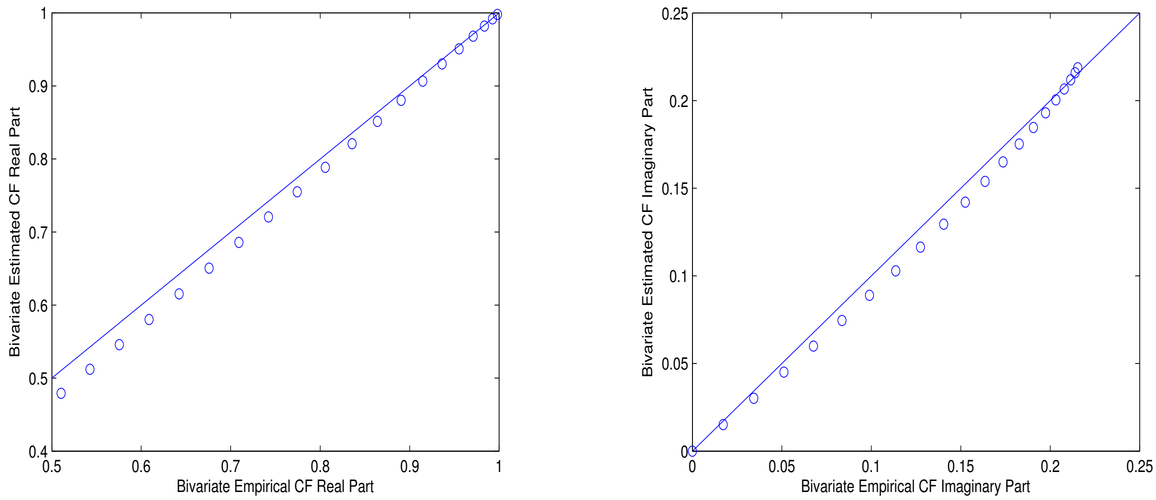
- $m = 1$

Figure 3.1: RSBS Univariate Characteristic Function Test



- $m = 2$

Figure 3.2: RSBS Bivariate Characteristic Function Test



The visual test can also be used to test the fit in each regime. To filter the state for each observation, we use the method proposed by Janczura and Weron (2014), which is combined with the EM algorithm.

Assume $K = 2$ and define equation (3) (Appendix 2.A) at the last iteration of the EM algorithm as $P(Y_j = k|w_N)$. Then we can tell that an observation is most probably coming from regime k if $P(Y_j = k|w_N) > 0.5$, where $k = 1, 2$. We refer to this approach as the filtering method and we use it when testing single regimes below.

We use Data 1 and the EM algorithm to estimate the model parameters, and the resulting estimates are presented in Table 3.2.

Table 3.2: Estimated Parameter Values

	μ_1	σ_1	μ_2	σ_2	p_{12}	p_{21}
True	0.156	0.110	-0.096	0.219	0.045	0.143
EM	0.134	0.114	-0.023	0.227	0.036	0.151

Figures 3.3 and 3.6, 3.4 and 3.7, 3.5 and 3.8 show the plots in both the univariate and bivariate cases based on the whole data set, the observations corresponding to regime 1

and the observations corresponding to regime 2. Figures 3.3, 3.4, 3.6 and 3.7 show few deviations from the reference lines. Figures 3.5 and 3.8 depict similar graphs for regime 2, and they show obvious deviations especially in the imaginary parts. From Table 3.2, we can see that μ_2 has apparent differences from the true value, and only 271 out of 2000 observations are filtered for regime 2. Those two facts can be the reasons for the deviations in Figures 3.5 and 3.8.

- $m = 1$

Figure 3.3: RSBS Univariate Characteristic Function Test (based on all observations)

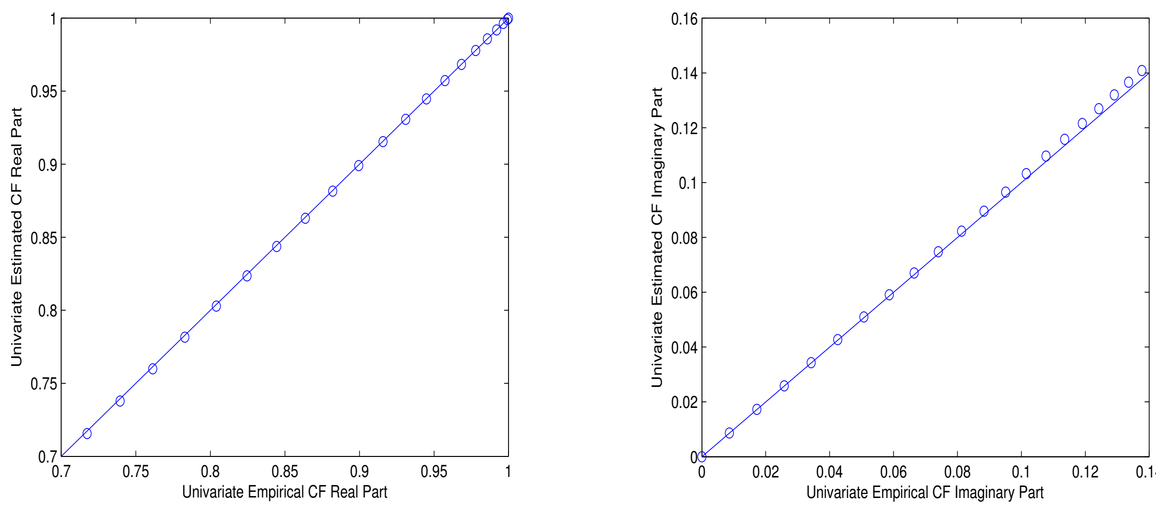


Figure 3.4: RSBS Regime1 Univariate Characteristic Function Test (based on observations corresponding to Regime 1)

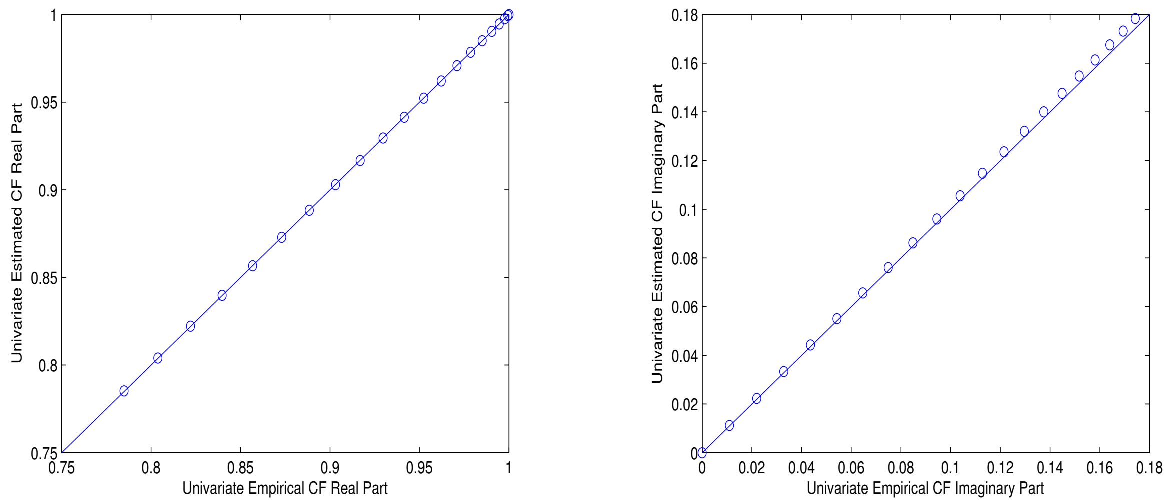
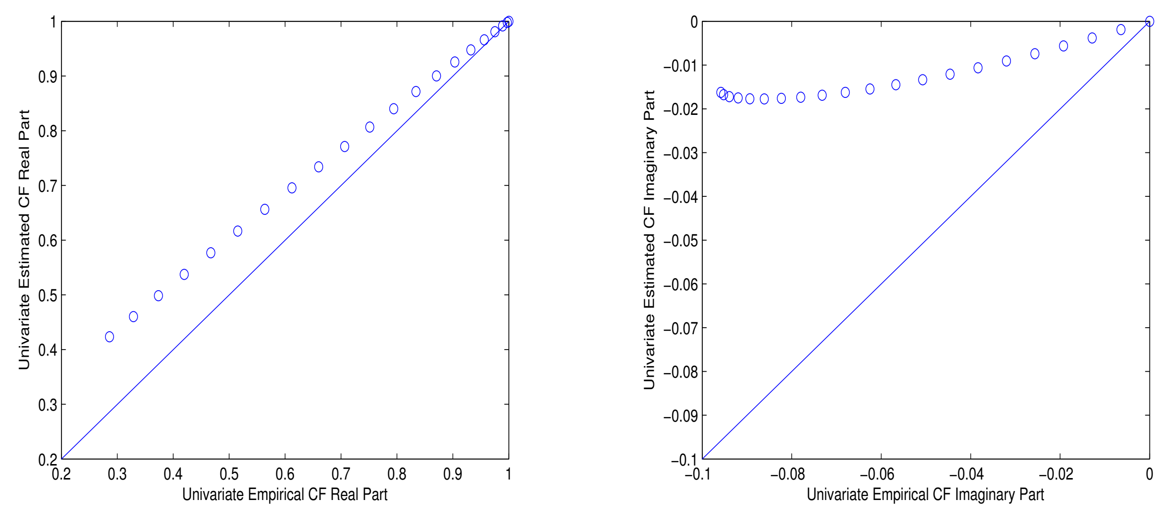


Figure 3.5: RSBS Regime2 Univariate Characteristic Function Test (based on observations corresponding to Regime 2)



- $m = 2$

Figure 3.6: RSBS Bivariate Characteristic Function Test (based on all observations)

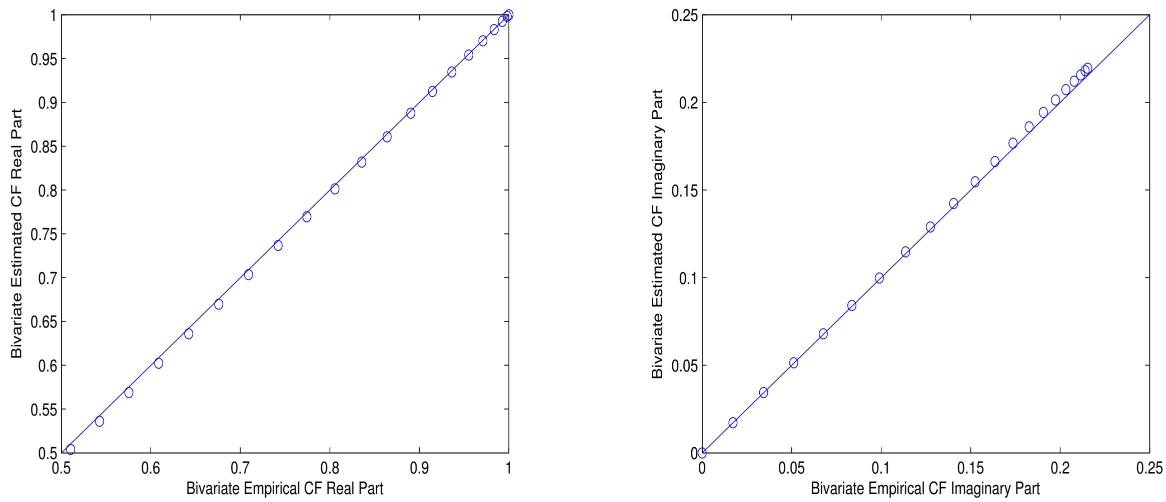


Figure 3.7: RSBS Regime1 Bivariate Characteristic Function Test (based on observations corresponding to Regime 1)

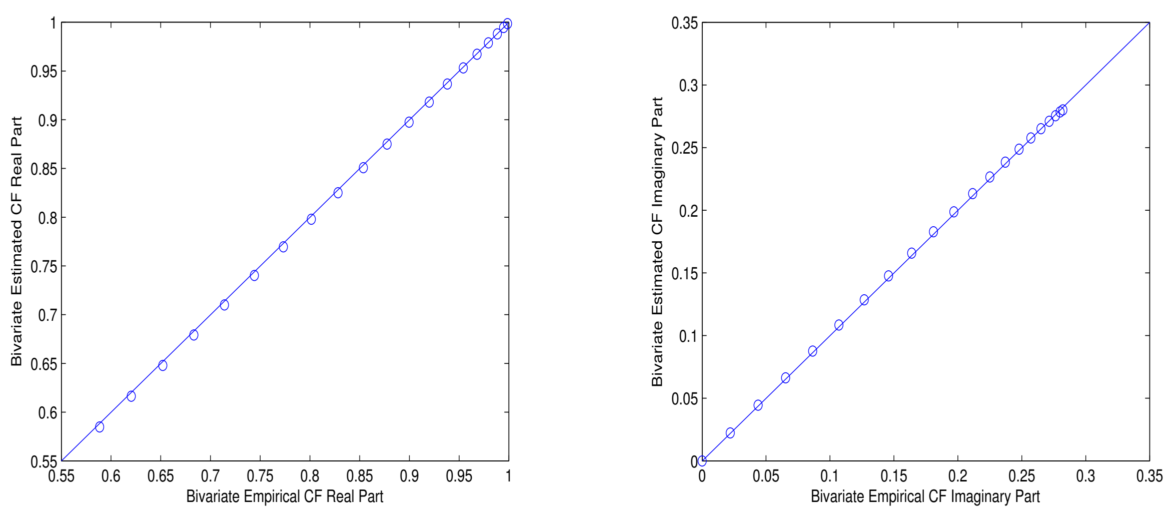
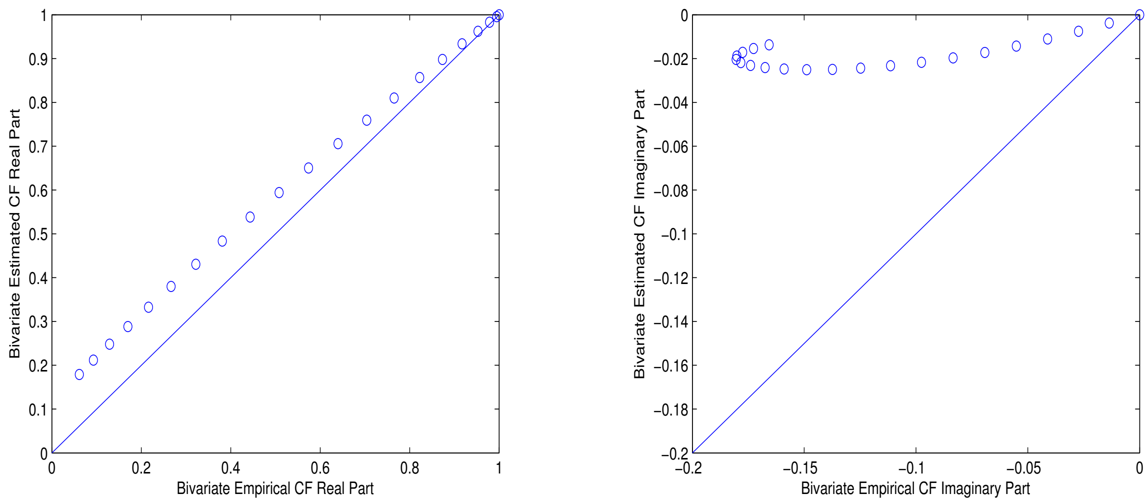


Figure 3.8: RSBS Regime2 Bivariate Characteristic Function Test (based on observations corresponding to Regime 2)



We have also applied the proposed visual test to the regime switching Black-Scholes model with other selections of model parameters and also to other regime switching models we specified in Section 1.1.2. Our results suggest that in some cases, the bivariate plots can give more information than the univariate ones do.

3.3.2 Method 2 – Statistical Test Based on the DECF Method

Here we focus on validating the proposed statistical test rather than testing the estimated model. We use simulated sets of data from the same regime switching Black-Scholes model as the one in the previous section. Before presenting the results, we first define the rejection percentage (RP), which represents a p-value as the proportion of the rejected null hypotheses based on N repetitions, that is,

$$RP := \frac{\#\left\{i : Q_N^i \geq Q_{\alpha,2q-p}\right\}}{N}, \ i = 1, ...N, \tag{3.9}$$

where Q_N^i is the test statistic at i^{th} simulation and α is the significance level for the test.

We assume $m = 1$ and still use the model parameters in Table 3.1. Other parameters are:

(i) $p = 0$ when model parameters are given, while $p = 6$ when model parameters are unknown and need to be estimated.

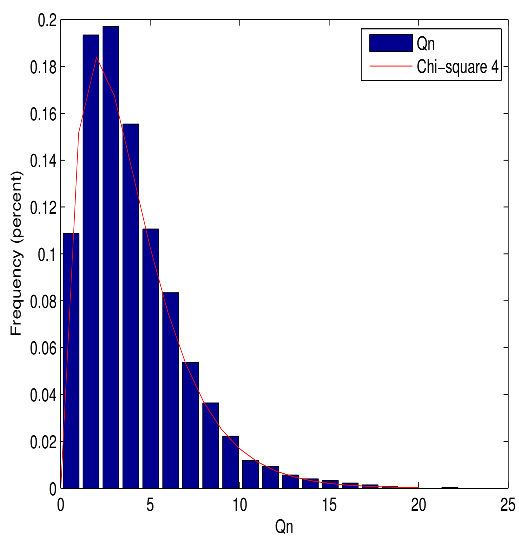
(ii) $N = 5000$ repetitions.

(iii) The significance level is $\alpha = 0.05$.

(iv) $q = 2$ when $p = 0$, and we will show further discussions of the q points when $p = 6$ below.

First we start with the case when model parameters are known. The $q = 2$ points obtained by the FDAWO method are $r_1 = 60$ and $r_2 = 100$. The theoretical test statistic is χ^2_4 in this case. In Figure 3.9, the thin line represents the density function of χ^2_4, while the histogram is based on 5000 simulated values of the test statistic given in equation (3.8). As the graph shows, they fit well, which confirms the asymptotic distribution of the test statistic. In addition, the rejection percentage in this case is 0.049, which is close to the significance level as desired.

Figure 3.9: Histogram of Simulated Test Statistics



Now we discuss the case when model parameters are unknown. This case is more complicated. For illustrations, we only consider $q = 4, 5, 6, 7$ and 8, and the corresponding degrees of freedom $2q - p$ are $2, 4, 6, 8$ and 10. Values of the q points are obtained by the FDAWO method as shown in Table 3.3.

Table 3.3: Selected Points for Statistical Test

Degree	Points
2	0.1 32.9 85.2 100
4	0.1 14.7 51.6 87.6 100
6	20.5 45.2 63.6 68.3 99.4 100
8	4.3 15.2 56.0 70.7 70.6 94.1 99.9
10	1.3 22.1 43.9 62.7 72.9 94.2 99.9 100

Using the above points, we simulate values of the test statistics for each of the degrees of freedom. In Table 3.4, we show the RP values for those different cases. We can see that the absolute difference between the RP and the significance level becomes smaller when the value of q becomes larger. These results suggest that a proper selection of the points is crucial for the accuracy of the asymptotic result stated in Theorem 3.2.2.

Table 3.4: Rejection Percentages for RSBS

Degree	2	4	6	8	10
RP	0.303	0.290	0.143	0.103	0.101

Next, we consider other selections of points, which are obtained by the quantization method described in Section 2.3.2. Tables 3.5 and 3.6 provide the selected sets of points, where G_N in (2.54) is respectively assumed to be a uniform distribution and the empirical characteristic function of observations. We refer to the former case as Quant1 and the latter one as Quant2.

Table 3.5: Selected Points by Quant 1 for Statistical Test

Degree	Points
2	12.5 62.5 37.5 87.5
4	10.0 70.0 50.0 90.0 30.0
6	8.3 75.0 41.6 91.7 24.9 58.3
8	7.1 78.4 35.5 92.8 21.3 64.0 49.7
10	6.5 69.4 32.1 93.9 19.4 57.0 44.6 81.7

Table 3.6: Selected Points by Quant 2 for Statistical Test

Degree	Points
2	12.2 48.3 29.4 73.7
4	10.3 56.7 39.6 79.3 24.8
6	8.7 62.1 33.0 83.1 20.9 46.2
8	7.4 66.5 28.7 86.0 18.2 51.9 39.7
10	6.5 55.5 25.0 87.6 16.1 44.2 34.2 69.2

Table 3.7 shows the rejection percentages for different cases. We find that Quant1 and Quant2 have similar [1] rejection percentages for each of the degrees of freedom. In addition, we need fewer degrees of freedom in those two cases than that in Table 3.4 for the absolute difference between the RP and the significance level to be less than 0.01. Therefore, we recommend using points selected by the quantization method to conduct this goodness-of-fit test when model parameters need to be estimated, where G_N in (2.54) can be either a uniform distribution or the empirical characteristic function of observations. In addition, the degrees of freedom should be at least equal to 10 in this example.

[1] They look almost identical, but they vary for different models and values of model parameters.

Table 3.7: Rejection Percentages for RSBS

Degree	2	4	6	8	10
Quant1	0.713	0.392	0.205	0.095	0.047
Quant2	0.718	0.392	0.205	0.095	0.047

We can summarize our finding as follows:

(i) The degrees of freedom at least at which the rejection percentage approaches the significance level depends on models and their parameters. In practice, we can conduct a pre-test (simulation study) by which we can decide the least degrees of freedom need to be used in the proposed goodness-of-fit test.

(ii) Since model parameters are estimated by the DECF estimation method with $m = 2$, so it is more reasonable to consider this test with $m = 2$. Because we suffer from the problem of selection of the points at which a characteristic function is evaluated for $m \geq 2$, so we left this for future research.

3.4 Conclusions of Chapter 3

In Chapter 3, we investigate goodness-of-fit testing methods based on characteristic functions. We propose two goodness-of-fit tests for regime switching models, which are extensions of the methods proposed by Altman (2004) and Koutrouvelis and Kellermeier (1981) respectively. We use univariate visual tests to examine the fit of the testing model to observations and bivariate visual tests to explore the dependence structure of consecutive observations. Moreover, we show the feasibility of combining the EM algorithm (Appendix 2.A) with the filtering method to test the goodness-of-fit in each regime.

We also propose a formal test and establish its asymptotic distribution with a degree of freedom that depends on the number of the points at which a characteristic function is evaluated and the number of parameters need to be estimated. Our numerical studies suggest that the asymptotic distribution provides a good approximation when model parameters are known. However, when parameters need to be estimated, a proper selection of points is crucial and is still an open question.

Chapter 4

Applications to Real Data

In this chapter, we apply the estimation and testing methods proposed in Chapters 2 and 3 to real data, where we use the same S&P 500 data set as the one used in Hartman and Groendyke (2013). We consider two-regime switching Black-Scholes (RSBS), variance gamma (RSVG) and Merton (RSM) models[1]. Section 4.1 provides the selected points used for the estimation and the estimation results for these regime switching models. Because the problem of selection of the grid points remains open for the statistical test proposed in Section 3.2.2, in Section 4.2 we only apply the proposed visual test to the estimated models. Section 4.3 concludes.

4.1 Estimation

The points used for the estimation are obtained by minimizing (2.54), where G_N is the empirical characteristic function of the data. Similarly as in Section 2.4, (2.54) is calculated by a numerical integration method with N_L^m points, where N_L is a set of equally spaced points and m represents the m-fold Cartesian product. The initial points are randomly generated by the Sobol sequence generator.

The notations are consistent with those used in Chapters 2 and 3. The parameters we use are:

[1] Although we do not know the true model of the data, we only use those three models for illustrations of our proposed estimation, point selection and goodness-of-fit methods. Definitely, we can choose other models as candidate models for the data.

(i) $m = 2$ [2], the dimension of the random variable $\mathbf{W_{j,m}}$.

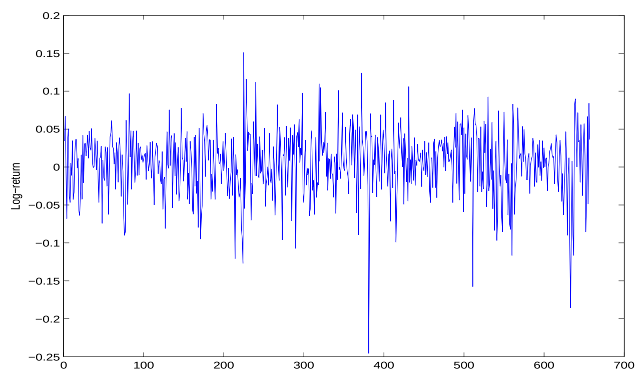
(ii) $q = 4, 5, 6$, the number of points used for the estimation of regime switching Black-Scholes, VG, and Merton models respectively .

(iii) The lower and upper bounds on the points are 0.1 and 100.

(iv) $N_L^2 = 25^2$, the number of points used in the numerical integration.

(v) The models are fitted to monthly total log-returns of the set of S&P 500 from February 1956 to October 2010 with total $T = 657$ observations. Figure 4.1 shows the log-returns of the data. The annual-averaged log-return is around 0.06.

Figure 4.1: The S&P 500 data plot



The resulting points used for the DECF estimation method are presented in Table 4.1.

[2]Based on our pre-analysis on the model parameters as described in Section 2.2.2, we choose $m = 2$ for our candidate models.

Table 4.1: Selected Points for Estimation

Model	r_1	r_2	r_3	r_4	r_5	r_6
RSBS	(16.3, 6.3)	(68.7, 69.7)	(16.6, 53.0)	(56.8, 9.1)		
RSVG	(16.4, 1.9)	(73.8, 65.1)	(21.4, 72.8)	(56.9, 8.8)	(15.0, 33.5)	
RSM	(16.9, 2.3)	(78.4, 59.8)	(12.5, 64.3)	(42.4, 81.2)	(16.8, 32.1)	(56.6, 7.1)

The estimated parameters [3] obtained by the DECF estimation method are as follows:

RSBS:

$$\mu_1 = 0.125, \quad \mu_2 = -0.142, \quad \sigma_1 = 0.110, \quad \sigma_2 = 0.230, \quad p_{12} = 0.046, \quad p_{21} = 0.148$$

RSVG

$$\mu_1 = 0.246, \quad \mu_2 = 0.477, \quad \sigma_1 = 0.121, \quad \sigma_2 = 0.172, \quad \nu_1 = 0.0001, \quad \nu_2 = 0.022,$$

$$\theta_1 = -0.023, \quad \theta_2 = -0.262, \quad p_{12} = 0.330, \quad p_{21} = 0.175.$$

RSM:

$$\mu_1 = 0.364, \quad \mu_2 = 0.272, \quad \sigma_1 = 0.100, \quad \sigma_2 = 0.114, \quad \mu_{J1} = 0.427, \quad \mu_{J2} = -0.087,$$

$$\lambda_1 = 0.244, \quad \lambda_2 = 0.329 \quad \sigma_{J1} = 0.471, \quad \sigma_{J2} = 0.065, \quad p_{12} = 0.487, \quad p_{21} = 0.169$$

The parameters for the RSBS model obtained by Hartman and Groendyke (2013) are:

$$\mu_1 = 0.156,\ \mu_2 = -0.096,\ \sigma_1 = 0.110,\ \sigma_2 = 0.219,\ p_{12} = 0.045,\ p_{21} = 0.143.$$

Table 4.2 compares the non-central moments of the set of S&P 500 data (Data) and the estimated RSBS models obtained by the DECF estimation method (DECF) and by Hartman and Groendyke (2013) (H&G). The moments of the data are obtained by the sample means and those of DECF and H&G are obtained based on fitted models. The numbers in the parentheses are the absolute differences between the moments based on the estimated models (DECF or H&G) and those of Data. We have several findings based on the results in Table 4.2:

(i) The second and forth order of the non-central moments are similar for the two estimated RSBS models, and the other two moments are very different, especially for the third moment.

[3] There are no standard errors accompanying the estimates, because these estimates are obtained by using the estimation method once.

(ii) The absolute differences between the first and second moments of DECF and those of Data are less than those between H&G and Data.

(iii) The absolute differences between the third and forth moments of DECF and those of Data are larger than those between H&G and Data.

(iv) The two cases, DECF and H&G, are similar in the accuracy of the estimation, since for both of them only two of the four non-central moments of the estimated model fit the data well.

Table 4.2: Comparison of non-central moments

Non-central moments (order)	1	2	3	4
Data	0.060	0.022	0.004	0.002
DECF	0.062 (0.002)	0.022(0.000)	-0.001 (0.005)	0.005(0.003)
H&G	0.096 (0.036)	0.021 (0.001)	0.004 (0.000)	0.004 (0.002)

4.2 Visual Test

Based on the proposed visual test in Section 3.2.1, we respectively plot the real and imaginary parts of $\bar{\Phi}_{m,N}(\cdot)$ versus those of $\hat{\Phi}_m(\cdot)$ for the two-regime switching Black-Scholes, Merton and variance gamma models as below. We consider both univariate ($m = 1$) and bivariate ($m = 2$) tests. Note that our goal is to test the goodness-of-fit of the estimated models respectively to the data, but is not to select the best model among all the candidate models.

- **Univariate Test** ($m = 1$)

  We use the same $q = 21$ points as those used in Section 3.3.1, $r_i = i - 1$, where $i = 1, ..., 21$, to conduct the univariate visual test.

  Figures 4.2-4.4 show the plots of the univariate real (left panels) and imaginary (right panels) parts of the empirical characteristic functions versus the estimated ones for the estimated RSBS, RSVG and RSM models. The solid straight lines in all the plots refer to the 45° line through the origin. We can see that only the plots of the RSBS model have few deviations from the reference line. Although, the plots of the real parts of the RSVG and RSM models are close to the reference line, the imaginary parts show obvious deviations. Therefore, the RSBS model has the best fit to the observations among all the candidate models, and we do not reject the RSBS model based on the univariate visual test.

Figure 4.2: RSBS Univariate Characteristic Functions Test for Whole Model

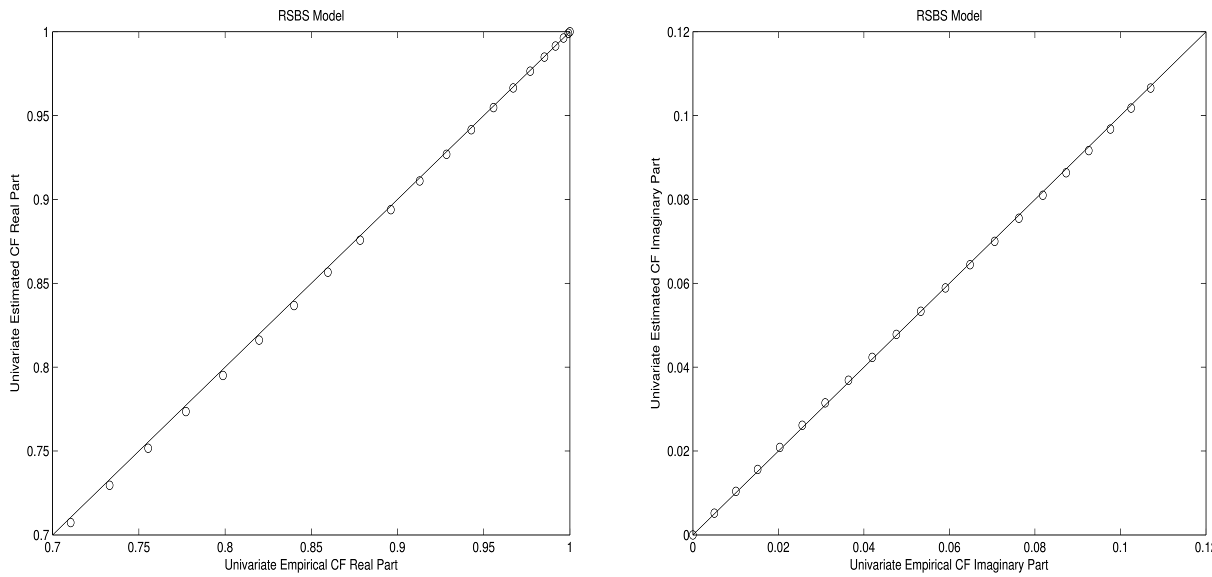


Figure 4.3: RSVG Univariate Characteristic Functions Test for Whole Model

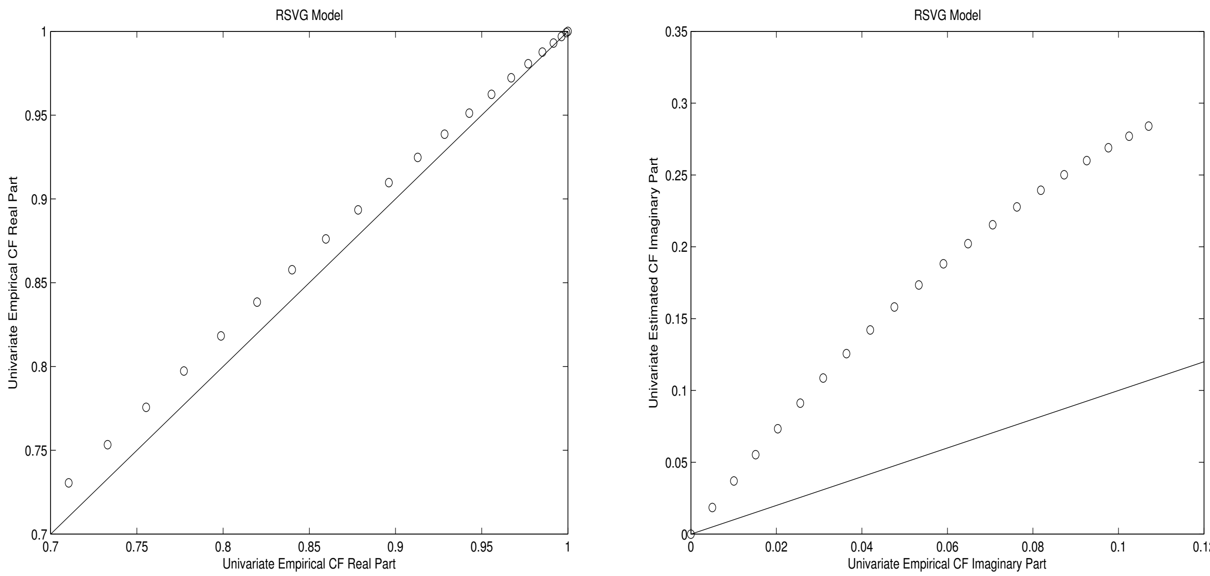
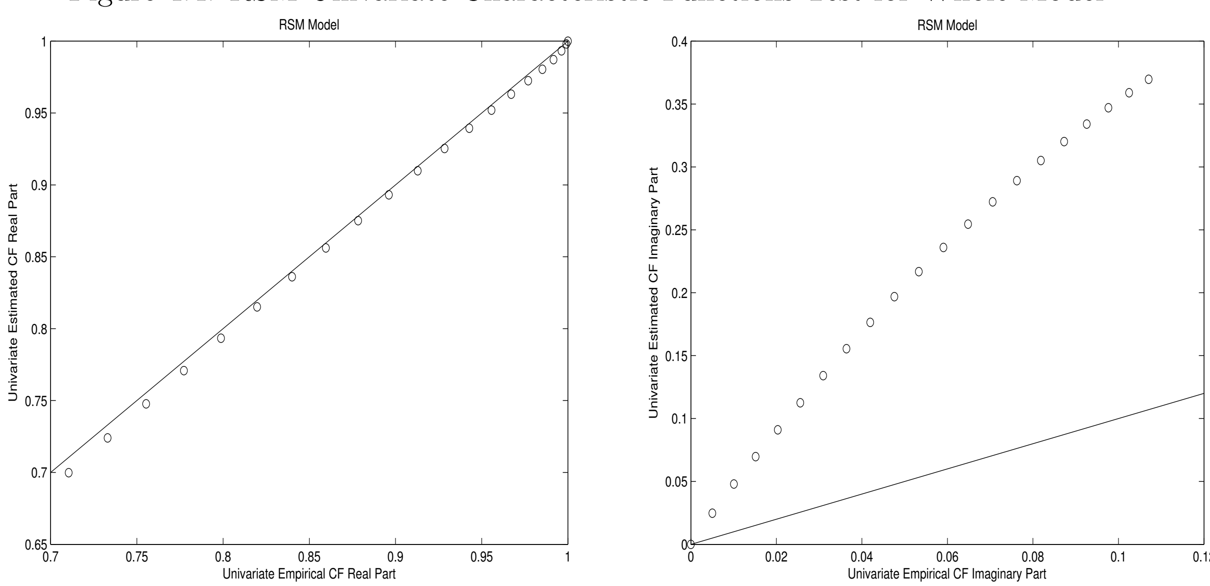


Figure 4.4: RSM Univariate Characteristic Functions Test for Whole Model



- **Bivariate Test** $(m = 2)$

To construct a bivariate visual test, we use the same $q = 21$ points as those used in Section 3.3.1, $r_i = (i-1, i-1)$, where $i = 1, ..., 21$, to conduct the bivariate visual test.

Figures 4.5-4.7 show the plots of the bivariate real (left panels) and imaginary (right panels) parts of the empirical characteristic functions versus the estimated ones for the estimated RSBS, RSVG and RSM models. We have similar conclusions as those reported in the univariate test above, except that the real parts of the RSVG and RSM models also show some deviations. Combining the results of the univariate and bivariate tests, we do not reject the RSBS model, but we reject the RSVG and RSM models.

Figure 4.5: RSBS Bivariate Characteristic Functions Test for Whole Model

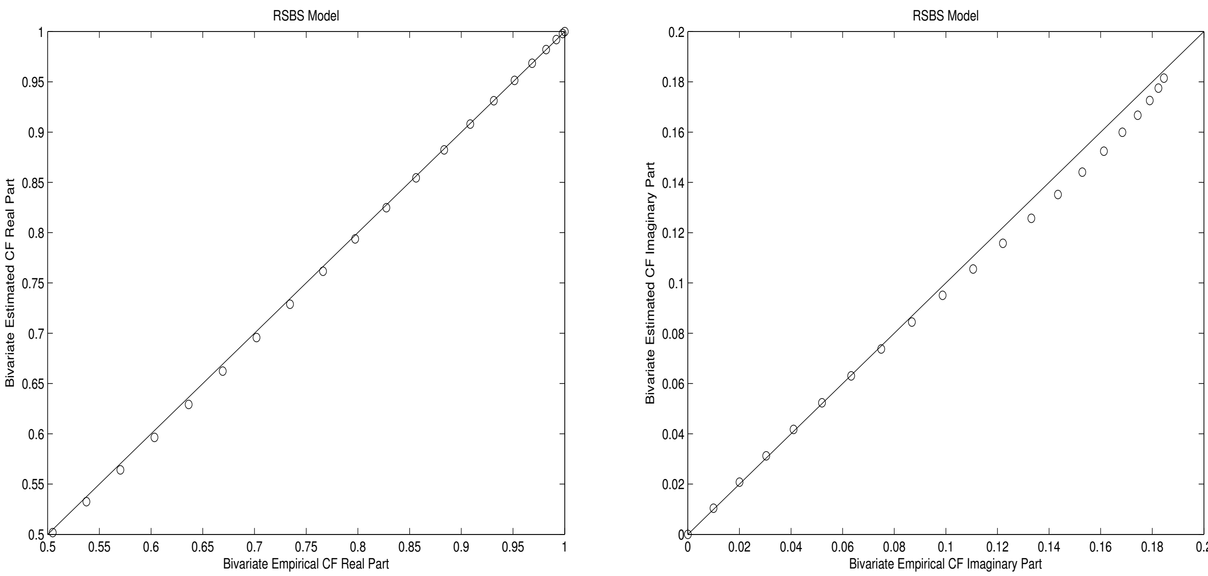


Figure 4.6: RSVG Bivariate Characteristic Functions Test for Whole Model

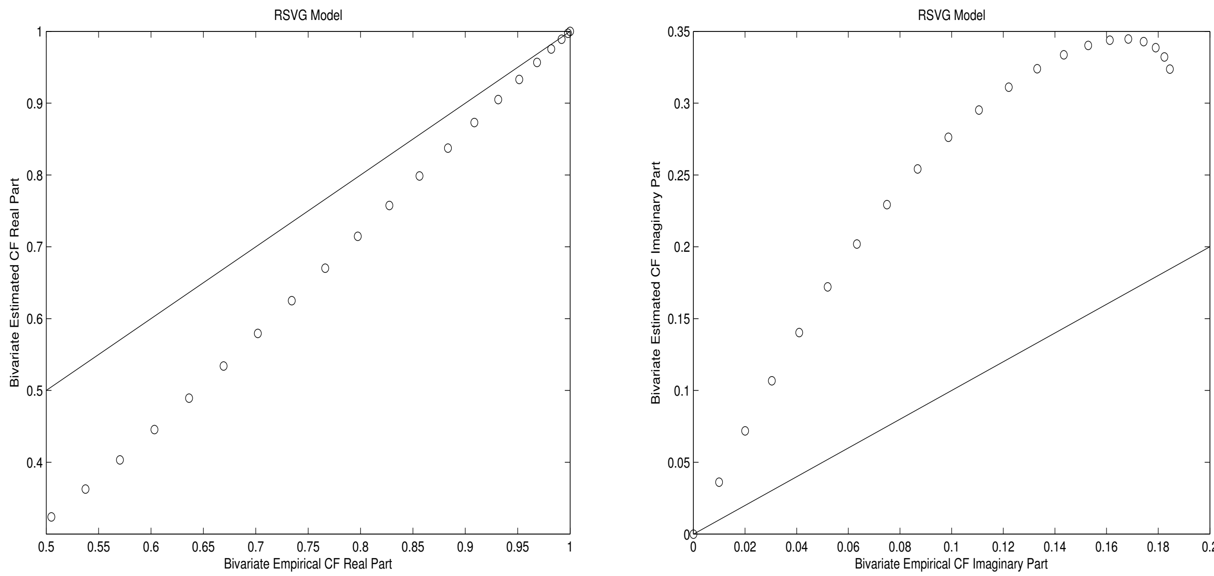
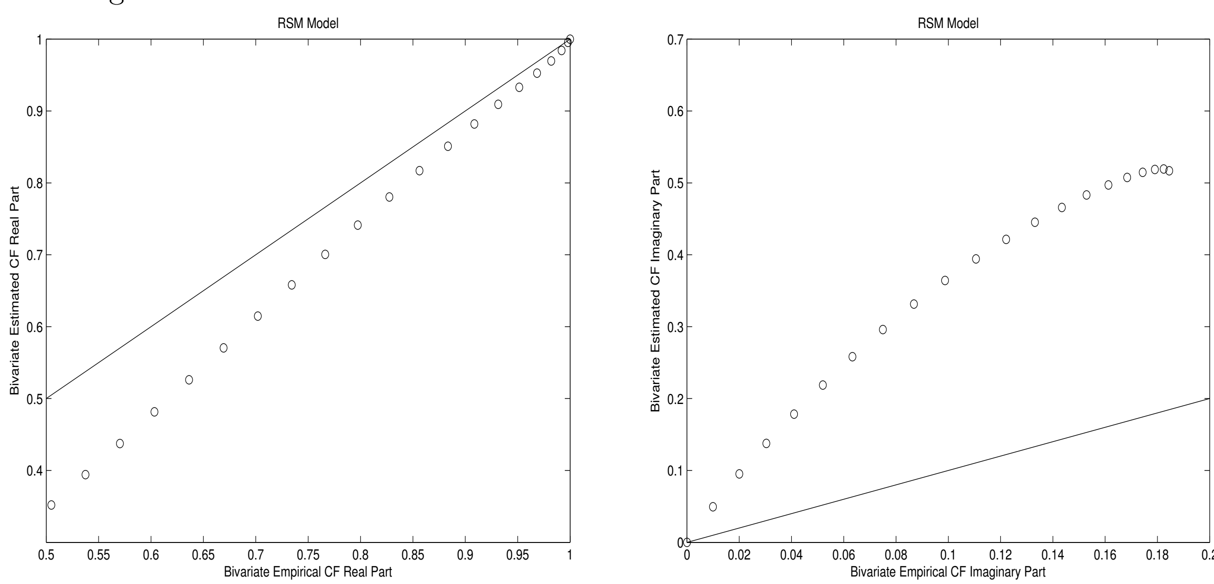


Figure 4.7: RSM Bivariate Characteristic Functions Test for Whole Model



4.3 Conclusions of Chapter 4

In this chapter, we apply the proposed DECF estimation method to estimate the two-regime switching Black-Scholes, Merton and variance gamma models, where we use a set of S&P 500 real data.

Based on the visual test results in Sections 4.2, we do not reject the estimated RSBS model, but reject the estimated RSVG and RSM models.

Note that there is a conflict between the rejection of the RSM model and the acceptance of the RSBS model, since the RSBS model is a special case of the RSM model with no jump parts. This may be explained by the following possible reasons:

(i) The initial values of parameters affect the estimation results. The DECF estimation method is not maximum-likelihood estimation, and the estimation procedure finds the closest model that fits observations. Therefore, the estimating function (2.8) with the estimated model parameters may only reach the local but not the global minimum.

(ii) The points used for these models in the estimation procedure are different and selected independently. Since they also affect the estimation results, we cannot say that the RSBS model is a nested model within the RSM model with respect to the estimation. To confirm this fact we have run a short simulation study. We estimate model parameters for both the RSBS model and the RSM model by using the same set of simulated data from an RSM model. Then we test the goodness-of-fit for both of the estimated models by using the proposed visual test. We repeat this procedure for different sets of simulated data (i.e., from the RSM model with different sets of parameters). We have found that for some of the cases the plots for the RSBS model show fewer deviations than those of the RSM model, which means we may infer the wrong model such that we do not reject the RSBS model but reject the RSM model based on the visual test. We can take this as a drawback of the visual test, which is informal and not robust.

Chapter 5

Pricing Ratchet Options under Regime Switching Models

5.1 Introduction and Motivation

In this chapter, we propose a numerical pricing method called the PV method to price Bermudan options, and we apply the PV method to ratchet equity-indexed annuities (EIAs) with early surrender risk under the model defined in (1.1).

EIAs are very popular contracts in insurance markets and gaining more attention in their proper pricing with variety features of the contracts and different assumptions of the models of the underlying fund. Kijima and Wong (2007) derive closed-form formulas for simple and compound ratchet EIAs under extended Vasicek interest rate models. Yuen and Yang (2010) consider the trinomial tree method and Wei et al. (2013) propose a lattice algorithm to price early surrender risk under CIR++ interest rate models.

Pricing ratchet EIAs with surrender risk is similar to pricing Bermudan options under certain assumptions. However, there are some challenges we need to handle. First, most contracts in insurance markets have longer maturities than financial options, and the models with a constant volatility or interest rate become questionable. Instead, regime switching models, where some economic factors or parameters are allowed to change randomly according to a Markov process, are more reasonable for pricing long-term financial and insurance products. Numerous authors have considered regime switching models because of their attractive features, including Hardy (2001, 2002), Yin et al. (2006), Lin et al. (2009), Yuen and Yang (2009), and Bastani et al. (2013). Second, the dimension of a model increases by introducing a Markov process, which makes pricing Bermudan options more difficult. Stochastic dynamic programming is a well-known technique that can be used to solve this problem. However, the recursive computation of the conditional expectations involving in the backward induction is also challenging in dynamic programming. Monte Carlo simulation is a natural approach to solve the high-dimensional problems when pricing Bermudan options since the publication of Tilley's paper (1993). Then the problem has been discussed by a considerable number of literature including Carriere (1996), Broadie and Glasserman (2004), Longstaff and Schwartz (2001), Andersen and Broadie (2004), Jin et al. (2007), Caramellino and Zanette (2011), and Boyle et al. (2001, 2013). For a more complete list of references, the reader may refer to Detemple (2006) and Hirsa (2013).

In this chapter, we present a general method of pricing Bermudan options for models where the characteristic functions of the underlying asset log-returns are known. We call this method the projected value (PV) method, where we also use the dynamic programming approach. However, we calculate the conditional expectations by representing the current value of the option with a series expansion and then applying the characteristic function. By doing this, we avoid Fourier inversions, which can be computationally intensive[1]. The idea of decomposing a payoff function using simpler functions is not new, since Bakshi and Madan (2000) and Chiarella et al. (1999) have considered this approach for European and American options respectively, where in the latter paper the authors use Hermite polynomials in the path-integral framework. Recently, Bang (2012) has proposed a pricing method based on characteristic functions for European vanilla options only, where he uses a trigonometric series to represent a payoff.

In Section 5.4, we compare the PV method with two well-known pricing methods. We show that the PV method can recover the COS method developed by Fang and Oosterlee (2008) if the basis functions in our approach are selected to be cosine series. The COS method has been applied to the problems of pricing Bermudan options under different distribution assumptions, and it is preferable to many alternative methods based on characteristic functions (see Fang and Oosterlee 2009, Ruijter and Oosterlee 2012). Although the points of departure for the PV method and the COS method are different, since the former uses cosine series expansions of the value functions while the latter uses cosine ex-

[1]None of the leading pricing methods based on characteristic functions use Fourier inversions, which often, due to oscillatory behaviour of the integrand, require a large number of integration points, and hence can be computationally intensive. We should mention that for numerical integration of inverse Fourier transform, the FFT method is not guaranteed in general to dominate a properly selected direct integration method. For an example, we refer to Chapter 4 in Zhu (2009). Some discussion of pricing methods based on inversions of characteristic functions can be found in Lord et al. (2008).

pansions of the density function, they produce the same pricing formula. However, the PV method allows for a much larger selection of approximating functions and hence is amenable to different strategies designed to enhance computational efficiency. This feature of the PV method is particularly useful for pricing ratchet options embedded in EIAs, for which the COS method exhibits low rate of convergence. To compare with the least-squares (LS) method proposed by Carrière (1996) and Longstaff and Schwartz (2001), the PV method has several advantages when pricing Bermudan options especially when the log-returns follow regime switching models in Lévy processes.

Other approaches to pricing Bermudan options based on Fourier methods include the Convolution (CONV) method (Lord et al. (2008)) and a related Fourier Space Time-stepping (FST) algorithm (Jackson et al. (2007)), where the latter utilizes the advantages of Fourier transform methods by transforming a partial integro-differential equation (PIDE) into Fourier space. The computational complexity of the CONV method is the same as that of the COS method [2], but typically the error for the latter diminishes at a higher rate (Fang et al. (2009)). In addition, the main assumption of the CONV method is that the transition density $f(y|x)$ of log-prices depends on x and y only via their difference, which we do not make in this chapter. The FST method, on the other hand, can incorporate regime-switching stock price behaviour by transforming a system of PIDEs into Fourier space (Jackson et al. (2007, 2008)). However, changes in regimes occur in continuous time, while in this chapter we assume that they follow a discrete-time Markov chain, which is consistent with some of the models proposed in the actuarial literature (e.g., Hardy (2001)).

In Section 5.5, we apply the PV method to the problems of pricing ratchet EIAs with surrender risk under regime switching models in Lévy processes. Although using Fourier transform methods to price options is a well established technique in finance and insurance (see Lord et al. (2008), Jackson et al. (2007, 2008), Dufresne et al. (2009), Eberlein (2014); and the books of Tankov (2003), Cherubini et al. (2010), and Hirsa (2013)), it is difficult to apply such methods to path-dependent options and options that depend on several risk factors. One notable feature of the PV method is that it provides a unified framework for different distributions of the log-returns of the underlying asset, since only one component of the algorithm needs to be replaced.

This chapter is organized as follows. We describe the models used for pricing in Section 5.2. In Section 5.3, we propose the pricing method, the PV method, and compare it with two known alternative approaches, the COS method and the least-squares method

[2] Here the computational complexity assumes that an arithmetic operation with an individual element has complexity $O(1)$. We discuss complexity of the COS method in Section 5.4.1.

in Section 5.4. In Section 5.5 we apply the method to price ratchet EIAs, followed by implementation results in Section 5.6. Section 5.7 concludes.

5.2 The Model

In this chapter, we introduce $\mathcal{S} = \{S_t : \ t \geq 0\}$ as the price process of the underlying asset and $S_j := S_{t_j}$, for a given set of equally-spaced time points $t_j := jh, \ j = 0, \ldots, N$. For pricing purposes, model (1.1) should be defined under a risk-neutral measure Q. Then it can be represented as

$$W_{j+1} := \ln \frac{S_{j+1}}{S_j} = \mu(Y_j) + Z_{j+1}\Big(\underline{\xi}(Y_j)\Big), \quad j = 0, \ldots, N-1, \tag{5.1}$$

where $\mathcal{W} = \{W_{j+1}, j = 0, 1, ..., N-1\}$ can be seen as the log-return process of the underlying asset.

Under some risk-neutral measure Q, the expected returns of model (5.1) must have a particular form in each regime so that the model is arbitrage-free. From the general option pricing theory, this condition will be satisfied if the discounted price process follows a martingale. In our case, this implies that under the pricing measure, we must have

$$E[e^{-mhr} S_m | \mathcal{F}_n] = e^{-nhr} S_n, \quad \text{for } m \geq n, \tag{5.2}$$

where r is a continuously compounded interest rate and $\mathcal{F}_n$ is the σ-field generated by the processes $\mathcal{S}$ and $\mathcal{Y}$ up to time t_n. Since (5.2) can be written as

$$E\Big[\frac{S_m}{S_n} | \mathcal{F}_n\Big] = e^{rh}, \quad \text{for } m = n+1,$$

then the left-hand side becomes

$$E\Big[e^{\mu_k + Z(\underline{\xi}_k)} | \mathcal{F}_n\Big] = E\Big[e^{Z(\underline{\xi}_k)} | \mathcal{F}_n\Big] e^{\mu_k} = E\Big[e^{i(-i)Z(\underline{\xi}_k)} | \mathcal{F}_n\Big] e^{\mu_k} = \Psi(-i; \underline{\xi}_k) e^{\mu_k}.$$

Therefore, (5.2) is equivalent to

$$\Psi(-i; \underline{\xi}_k) e^{\mu_k} = e^{rh}, \quad \text{for} \quad k = 1, \ldots, K,$$

which leads to the condition

$$\mu_k = rh - \psi(-i; \underline{\xi}_k), \tag{5.3}$$

where $\psi(\cdot;\underline{\xi}_k)$ is the natural logarithm of the characteristic function of $Z(\underline{\xi}_k)$, that is, $\psi(\cdot;\underline{\xi}_k) := \log(\Psi(\cdot;\underline{\xi}_k))$. Define $M_Z(\cdot,\underline{\xi})$ as the moment generating function of $Z(\underline{\xi})$ and $m_Z(\cdot;\underline{\xi})$ as the cumulant moment generating function of $Z(\underline{\xi})$. By definitions,

$$M_Z(z,\underline{\xi}) := E(e^{zZ(\underline{\xi})}) \text{ and } m_Z(z,\underline{\xi}) := \log\Big(M_Z(z,\underline{\xi})\Big), z \in \mathbb{R}.$$

Since $\Psi(-i;\underline{\xi}_k) = M_Z\Big(i(-i),\underline{\xi}\Big) = M_Z(1,\underline{\xi})$ and similarly $\psi(-i;\underline{\xi}_k) = m_Z\Big(i(-i),\underline{\xi}\Big) = m_Z(1,\underline{\xi})$, then (5.3) can be rewritten as

$$\mu_k = rh - m_Z(1;\underline{\xi}_k), \tag{5.4}$$

where we assume that $m_Z(1;\underline{\xi}_k),\ k = 1,\ldots,K$, are finite. The pricing measure that we use in our numerical study is equivalent to equation (5.3), but the proposed pricing method is valid for any selection of risk-neutral measure Q[3].

We also consider an extension of model (5.1) where the timing of the regime that determines the parameter $\underline{\xi}$ is different. In model (5.1), we assume that $\underline{\xi}$ at the next time period t_{j+1} depends on the regime Y_j at time t_j, and the variable Z_{j+1} is the only uncertain term in the log-return $\mu(Y_j) + Z_{j+1}(\underline{\xi}(Y_j))$. An alternative model assumes that $\underline{\xi}$ at time t_{j+1} also depends on the regime realization at t_{j+1}, then model (5.1) becomes

$$W_{j+1} := \ln\frac{S_{j+1}}{S_j} = \mu(Y_j) + Z_{j+1}\Big(\underline{\xi}(Y_{j+1})\Big), \quad j = 0,\ldots,N-1. \tag{5.5}$$

Similar specifications have been considered in the context of discrete-time stochastic volatility models, and we refer to Durham (2006) for a discussion of pros and cons of different formulations of such models. In the original model (5.1), the current regime Y_j determines uniquely all of the parameters in the distribution of the return W_{j+1} over the interval $[t_j, t_{j+1}]$, whereas the return follows a mixture distribution in (5.5).

By conditioning, the characteristic function of the log-return W_{j+1} in model (5.5) can be expressed in terms of the characteristic function of $Z(\underline{\xi})$, which has the form of

$$E\Big[e^{iz\big(\mu(Y_j)+Z_{j+1}(\underline{\xi}(Y_{j+1}))\big)}|Y_j = k\Big] = e^{iz\mu(k)}\sum_{l=1}^{K}\Psi(z;\underline{\xi}_l)p_{kl}. \tag{5.6}$$

To ensure that the discounted price process forms a martingale under model (5.5), in each regime the parameters must satisfy

$$\mu(k) = rh - \ln\left(\sum_{l=1}^{K}\psi(-i;\underline{\xi}_l)p_{kl}\right). \tag{5.7}$$

[3]For example, Lin et al. (2009) use the Esscher transform to determine an equivalent pricing measure.

The proposed estimation method can be adjusted for this extended model (5.5).

5.3 The Pricing Method

In this section, we present a numerical method of pricing Bermudan options.

Assume the payoff function of an option is $G(\tau, S_\tau, Y_\tau)$ at the exercise time τ, where $\tau \in \mathcal{T} := \{t_0, t_1, \ldots, t_N = T\}$, that is, the option can be exercised prior to maturity at $N+1$ time points including the initial time. Our objective is to find the value

$$V := \max_{\tau \in \mathcal{T}} E\Big[e^{-\tau r} G(\tau, S_\tau, Y_\tau)\Big],$$

where the maximum is over all possible stopping times $\{0, 1, \ldots, N\}$, and the expectation is taken under a risk-neutral pricing measure Q.

By the dynamic programming principle, we can obtain V by calculating $V(t_j, \cdot, \cdot)$ through the following backward recursive algorithm:

$$V(T, s, y) = G(T, s, y), \tag{5.8}$$

$$V(t_j, s, y) = \max\Big[G(t_j, s, y),\ C(t_j, s, y)\Big], \quad j = N-1, ..., 0, \tag{5.9}$$

where the continuation value, $C(t_j, s, y)$, is defined as

$$C(t_j, S, y) := e^{-rh} E\Big[V(t_{j+1}, S_{t_{j+1}}, Y_{t_{j+1}}) | S_{t_j} = s, Y_{t_j} = y\Big]. \tag{5.10}$$

Then, the price of the option V can be obtained by taking $V = V(0, S_0, Y_0)$. In practice, we must calculate or accurately approximate the conditional expectations in (5.10) for all regimes $y \in \mathbf{S}$ and some selected points s from the state space of the price process $\mathcal{S}$. Recursively computing these expectations with the dynamic programming principle is another challenge. Some existing numerical methods for the dynamic programming, including crude Monte-Carlo simulation, least squares regression and mesh point method, have been proposed in the existing literature.

In this chapter, we propose using the characteristic functions of the log-increments of the process $\mathcal{S}$ to approximate the conditional expectations (5.10) in the above dynamic programming setup. Define $\hat{V}(t_{j+1}, \cdot)$ as the approximation of $V(t_{j+1}, \cdot)$ at time t_{j+1}, $j = 0, \ldots, N-1$, obtained from the backward recursion. In addition, let $\mathcal{M} := \{m_l(x) : x \in \mathcal{X}, l = 0, .., L\}$ be a collection of basis functions defined on $\mathcal{X} = \{x : \exp(x) \in \mathcal{S}\}$. Then the approximation $\hat{C}(t_j, \cdot, \cdot)$ of the continuation value $C(t_j, \cdot, \cdot)$ at time t_j is obtained in the following two steps:

(S4–1) For each regime $y \in \mathcal{Y}$, approximate the function $\hat{V}^e(t_{j+1}, \cdot, y) := \hat{V}(t_{j+1}, \exp(\cdot), y)$ by a weighted sum of the basis functions

$$\hat{V}^e(t_{j+1}, x, y) \approx \hat{V}^e_{\mathcal{M}}(t_{j+1}, x, y) := \sum_{l=0}^{L} \alpha_l(y) m_l(x), \tag{5.11}$$

where $\alpha_l(\cdot)$ are the assigned weights or coefficients of the basis functions $m_l(\cdot),\ l = 0, ..., L$.

(S4–2) Calculate the conditional expectation of $\hat{V}^e_{\mathcal{M}}$ as follows:

$$\begin{aligned} E\Big[\hat{V}^e_{\mathcal{M}}(t_{j+1}, \ln(S_{t_{j+1}}), Y_{t_{j+1}}) | S_{t_j} = s, Y_{t_j} = y\Big] \\ &= \sum_{l=0}^{L} E\Big[\alpha_l(Y_{t_{j+1}}) m_l(\ln(S_{t_{j+1}})) | S_{t_j} = s, Y_{t_j} = y\Big] \\ &= \sum_{l=0}^{L} E\Big[\alpha_l(Y_{t_{j+1}}) | Y_{t_j} = y\Big] E\Big[m_l\left(\ln(S_{t_{j+1}})\right) | S_{t_j} = s, Y_{t_j} = y\Big], \end{aligned} \tag{5.12}$$

where the last equation results from the assumption that the Markov chain $\mathcal{Y}$ is independent of the price process $\mathcal{S}$. By the representation (5.1), the variables $S_{t_{j+1}}$ and $Y_{t_{j+1}}$ are independent given S_{t_j} and Y_{t_j}.

We should note that in step (S4–1), we approximate the current value of the option $\hat{V}$ on a logarithmic scale, which makes computations conveniently for step (S4–2). We can also obtain the approximation by any of the standard techniques, like the least-squares method or an interpolation technique. For example, the coefficients $\alpha_l(y) \equiv \alpha_l(t_{j+1}, y),\ l = 0, \dots, L$, may solve the following optimization problem

$$(\alpha_0(y), \dots, \alpha_L(y)) := \arg \inf_{\beta_0, \dots, \beta_L} \sum_{i=1}^{M} \left(\hat{V}^e_L(t_{j+1}, x_i^{j+1}, y) - \sum_{l=0}^{L} \beta_l m_l(x_i^{j+1}) \right)^2, \tag{5.13}$$

where the points $\{x_1^{j+1}, \dots, x_M^{j+1}\}$ are sampled randomly from a pre-specified probability distribution on $\mathcal{X}$.

In this chapter, we focus on trigonometric functions, which is motivated by the fact that steps (S4–1)–(S4–2) are easy to implement for these functions. Based on classical Fourier analysis, properly scaled trigonometric functions form a complete and orthogonal set in the space of square integrable functions over any given interval $[-l, l]$, $l > 0$. In particular, any square-integrable function with a period $2l$ on $\mathbb{R}$ can be approximated arbitrarily close in the L^2-sense by sums of trigonometric functions. In step (S4–1), $\hat{V}^e(t_{j+1}, \cdot, y)$ is typically non-periodic. In order to represent it as a sum of trigonometric functions, we first need to truncate the range of possible values of $\ln(S_{t_{j+1}})$ to a finite interval, for example, $(-l_X, l_X)$, where $l_X \equiv l(t_{j+1})$ is a suitably chosen constant. Then, we can approximate $\hat{V}^e(t_{j+1}, \cdot, y)$ by its orthogonal projection onto the space spanned by the following set of functions

$$\mathcal{M}_{(l_X,L)} := \Big\{1, \cos(\frac{\pi x}{l_X}), \ldots, \cos(\frac{\pi L x}{l_X}), \sin(\frac{\pi x}{l_X}), \ldots, \sin(\frac{\pi L x}{l_X});\ x \in [-l_X, l_X]\Big\}. \tag{5.14}$$

Then $\hat{V}^e(t_{j+1}, \cdot, y)$ can be approximated by

$$\hat{V}^e_{\mathcal{M}_{(l,L)}}(t_{j+1}, x, y) := A(y) + \sum_{l=1}^{L} \Big\{a_l(y)\cos(\frac{\pi l x}{l_X}) + b_l(y)\sin(\frac{\pi l x}{l_X})\Big\}, \qquad x \in (-l_X, l_X), \tag{5.15}$$

where $A(y) = a_0(y)/2$,

$$a_l(y) \;:=\; \frac{1}{l_X}\int_{-l_X}^{l_X} \hat{V}^e(t_{j+1}, x, y)\cos(\frac{\pi l x}{l_X})dx, \quad l = 0, 1, 2, \ldots, L, \tag{5.16}$$

and

$$b_l(y) \;:=\; \frac{1}{l_X}\int_{-l_X}^{l_X} \hat{V}^e(t_{j+1}, x, y)\sin(\frac{\pi l x}{l_X})dx, \quad l = 1, 2, \ldots, L. \tag{5.17}$$

Typically, we need to approximate the coefficients a_l, $l = 0, 1, \ldots, L$, and b_l, $l = 1, \ldots, L$, defined in (5.16)–(5.17) by some numerical methods, or we can obtain these coefficients by using a fast Fourier transform (FFT).

To calculate the conditional expectation of $\hat{V}^e_{\mathcal{M}}$ in step (S4–2), we need to find two sets of expectations $E[\alpha_l(Y_{t_{j+1}})|Y_{t_j} = y]$, $l = 0, \ldots, L$, and $E[m_l(\ln(S_{t_{j+1}}))|S_{t_j} = s, Y_{t_j} = y]$, $l = 0, \ldots, L$. The first expectations depend only on the transition probabilities of the Markov chain $\mathcal{Y}$, and they can be calculated by a weighted sum of the given coefficients $\alpha_l(y)$ as:

$$E\Big[\alpha_l(Y_{t_{j+1}})|Y_{t_j} = k\Big] = \sum_{i \in K} \alpha_l(y_i) p_{ki}, \quad l = 0, \ldots, L.$$

Calculating the second set of expectations is challenging, for which a proper selection of basis functions is crucial. We may impose two requirements on the basis functions $\mathcal{M}$ to make the calculation easier:

(A3–1) For each function $m_l \in \mathcal{M}$, the conditional expectation

$$E\Big[m_l\left(\ln(S_{t_{j+1}})\right)|S_{t_j}=s, Y_{t_j}=y\Big] \tag{5.18}$$

can be represented in terms of the characteristic function Ψ.

(A3–2) For each regime y, the conditional expectation (5.18) can be represented as an explicit function of $x \equiv \ln(s) \in \mathcal{X}$.

Indeed, if (A3–1) is satisfied, typically the expectation (5.18) can be expressed in an analytical form. If (A3–2) is also satisfied, then calculating the conditional expectations in the recursive backward procedure will be convenient.

For model (5.1), polynomials and trigonometric functions satisfy the two requirements (A3–1) and (A3–2).

Polynomials:

When $m_l(x) := x^l,\ l \geq 0$, is a monomial function, we can use the binomial theorem to show that

$$\begin{aligned}
& E\Big[(\ln S_{t_{j+1}})^l|S_{t_j}=s, Y_{t_j}=k\Big] = E\Big[(\ln S_{t_j}+W_{t_{j+1}})^l|S_{t_j}=s, Y_{t_j}=k\Big] \\
= \;& E\Big[(\ln s+W_{t_{j+1}})^l|Y_{t_j}=k\Big] = \sum_{m=0}^{l}\binom{l}{m}(\ln s)^{l-m}E\Big[(W_{t_{j+1}})^m|Y_{t_j}=k\Big] \\
= \;& \sum_{m=0}^{l}\binom{l}{m}(\ln s)^{l-m}\frac{1}{i^m}\frac{\partial^m\Psi_k(z)}{\partial z^m}\Big|_{z=0},
\end{aligned} \tag{5.19}$$

where $\Psi_k(z)=\exp(iz\mu_k)\Psi(z;\underline{\xi}_k)$ is the characteristic function of $\mu_k+Z(\underline{\xi}_k)$. Then we can implement (S4–1) and (S4–2) efficiently if all the required derivatives of $\Psi_k(z)$ can be obtained.

Note that for some cases of model (5.1), we can express the conditional moments of $S_{t_{j+1}}$ directly in terms of its characteristic function or moment generating function without taking the logarithmic transformation; however, this approach is limited because increments of Lévy processes typically have exponential tails and only a small number of moments of $S_{t_{j+1}}$ may exist.

Trigonometric:

When $m_l(x) := \cos(\omega x)$ and/or $\sin(\omega x)$, we can use elementary properties of trigonometric functions to show that for any real number ω

$$\begin{aligned} & E\Big[\cos\left(\omega \ln(S_{t_{j+1}})\right)|S_{t_j}=s, Y_{t_j}=k\Big] = E\Big[\cos\Big(\omega\left(\ln(S_{t_j})+W_{t_{j+1}}\right)\Big)|S_{t_j}=s, Y_{t_j}=k\Big] \\ = & \cos\Big(\omega \ln(s)\Big)E\Big[\cos(\omega W_{t_{j+1}})|Y_{t_j}=k\Big] - \sin\Big(\omega\ln(s)\Big)E\Big[\sin(\omega W_{t_{j+1}})|Y_{t_j}=k\Big] \\ = & \cos\Big(\omega\ln(s)\Big)\Re\Big(\Psi_k(\omega)\Big) - \sin\Big(\omega\ln(s)\Big)\Im\Big(\Psi_k(\omega)\Big), \end{aligned} \tag{5.20}$$

where $\Re(z)$ and $\Im(z)$ denote the real part and the imaginary part of a complex number z respectively.

Similarly, we have

$$E\Big[\sin\left(\omega\ln(S_{t_{j+1}})\right)|S_{t_j}=s, Y_{t_j}=k\Big] = \cos\Big(\omega\ln(s)\Big)\Im\Big(\Psi_k(\omega)\Big) + \sin\Big(\omega\ln(s)\Big)\Re\Big(\Psi_k(\omega)\Big). \tag{5.21}$$

Formulas (5.20) and (5.21) show that the requirements (A3–1) and (A3–2) are satisfied for the basis functions coming from $\mathcal{M}_{(l_X,L)}$. Calculating the conditional expectations is equivalent to evaluating the real and imaginary parts of the characteristic function $\Psi(\cdot;\underline{\xi}_k)$ at different points. Therefore, we can find the expression for the conditional expectation $\hat{V}^e_{\mathcal{M}_{(l_X,L)}}(t_{j+1},\ln(S_{t_{j+1}}),Y_{t_{j+1}})$ based on equations (5.20) and (5.21) :

$$\begin{aligned} E\Big[\hat{V}^e_{\mathcal{M}_{(l_X,L)}}\Big(t_{j+1},\ln(S_{t_{j+1}}),Y_{t_{j+1}}\Big)|S_{t_j}=s,Y_{t_j}=k\Big] \quad &= \quad \frac{1}{2}E\Big[a_0(Y_{t_{j+1}})|Y_{t_j}=k\Big] \\ + \sum_{l=1}^{L}\Big[\bar{a}_l(k)\cos\Big(\frac{\pi l}{l_X}\ln(s)\Big) + \bar{b}_l(k)\sin\Big(\frac{\pi l}{l_X}\ln(s)\Big)\Big]R_l(k) \\ + \sum_{l=1}^{L}\Big[\bar{b}_l(k)\cos\Big(\frac{\pi l}{l_X}\ln(s)\Big) - \bar{a}_l(k)\sin\Big(\frac{\pi l}{l_X}\ln(s)\Big)\Big]I_l(k), \end{aligned} \tag{5.22}$$

with

$$\bar{a}_l(k) := E\Big[a_l(Y_{t_{j+1}})|Y_{t_j}=k\Big], \qquad \bar{b}_l(k) := E\Big[b_l(Y_{t_{j+1}})|Y_{t_j}=k\Big], \quad l=1,2,\ldots,L, \tag{5.23}$$

and

$$R_l(k) := \Re\Big(\Psi_k(\frac{\pi l}{l_X})\Big), \qquad I_l(k) := \Im\Big(\Psi_k(\frac{\pi l}{l_X})\Big), \qquad l=1,2,\ldots,L. \tag{5.24}$$

We should note that (5.22) can be written as

$$
\begin{aligned}
E\Big[\hat{V}^e_{\mathcal{M}_{(l_X,L)}}\Big(t_{j+1}, \ln(S_{t_{j+1}}), Y_{t_{j+1}}\Big)|S_{t_j}=s, Y_{t_j}=k\Big] &= \frac{1}{2}E\Big[a_0(Y_{t_{j+1}})|Y_{t_j}=k\Big] \\
&+ \sum_{l=1}^{L}\Big[\bar{a}_l(k)R_l(k)+\bar{b}_l(k)I_l(k)\Big]\cos\Big(\frac{\pi l}{l_X}\ln(s)\Big) \\
&+ \sum_{l=1}^{L}\Big[\bar{b}_l(k)R_l(k)-\bar{a}_l(k)I_l(k)\Big]\sin\Big(\frac{\pi l}{l_X}\ln(s)\Big). \qquad (5.25)
\end{aligned}
$$

Therefore, the conditional expectation has an analytical form. The coefficients for $\cos(\pi l\ln(s)/l_X)$ and $\sin(\pi l\ln(s)/l_X)$ in equation (5.25) do not depend on s, so they can be calculated only once at each time period and in each regime.

In addition, the PV method can be modified for the alternative model (5.5). For any basis function, we have

$$
\begin{aligned}
&E\Big[m_l\Big(\ln(S_{t_{j+1}})\Big)|S_{t_j}=s, Y_{t_j}=y_{k_1}, Y_{t_{j+1}}=y_{k_2}\Big] \\
&\quad = E\Big[m_l\Big(\ln(S_{t_j})+\mu(Y_{t_j})+Z_{t_{j+1}}\Big(\underline{\xi}(Y_{t_{j+1}})\Big)\Big)|S_{t_j}=s, Y_{t_j}=y_{k_1}, Y_{t_{j+1}}=y_{k_2}\Big] \\
&\quad = E\Big[m_l\Big(\ln(s)+\mu(y_{k_1})+Z_{t_{j+1}}(y_{k_2})\Big)|S_{t_j}=s, Y_{t_j}=y_{k_1}, Y_{t_{j+1}}=y_{k_2}\Big] \\
&\quad = E\Big[m_l\Big(\ln(s)+\mu(y_{k_1})+Z_{t_{j+1}}(y_{k_2})\Big)\Big],
\end{aligned}
$$

for which the last line is based on the assumption of the conditional independence of $Z_{t_{j+1}}$ on the remaining variables. conditional on $Y_{t_{j+1}}$, step (S4–2) in the PV method should be replaced with

(S4–2*) Calculate the conditional expectation of $\hat{V}^e_{\mathcal{M}}$ as follows:

$$
\begin{aligned}
&E\Big[\hat{V}^e_{\mathcal{M}}\Big(t_{j+1}, \ln(S_{t_{j+1}}), Y_{t_{j+1}}\Big)|S_{t_j}=s, Y_{t_j}=k\Big] \\
&\qquad = \sum_{l=0}^{L}\sum_{i=1}^{K}\alpha_l(y_i)E\Big[m_l\Big(\ln(s)+\mu(y_k)+Z_{t_{j+1}}(y_i)\Big)\Big]p_{ki}. \qquad (5.26)
\end{aligned}
$$

For the set $\mathcal{M}_{(l_X,L)}$, the conditional expectation in (5.26) can be expressed in terms of the characteristic function of $Z_{t_{j+1}}$ similarly as in (5.25), thus the PV method can be also applied to the alternative model (5.5).

5.4 Comparison with Other Methods

In Section 5.4.1, we show that we can recover the COS method proposed by Fang and Oosterlee (2008, 2009) by the PV method with a particular selection of basis functions. This result enables us to construct an improved version of the latter method, which is particularly useful for pricing ratchet options. In Section 5.4.2, we compare the PV approach with the least-squares method proposed by Carrière (1996) and Longstaff and Schwartz (2001). Since the original formulations of the COS and the LS method do not allow for regime changes, to simplify our exposition we assume that the Markov chain $\mathcal{Y}$ stays in one regime only.

5.4.1 The COS method

To present the COS method, suppose that at time t_j we want to find an approximation of the expectation

$$I(x_0) := \int U(x) f(x; x_0) dx, \tag{5.27}$$

where $U(x) \equiv \hat{V}^e(t_{j+1}, x, y)$ with a fixed value of y, $f(x; x_0)$ is the density function of $\ln(S_{t_{j+1}})$ given $\ln(S_{t_j}) = x_0$, and $\hat{V}^e$ is the value function in the log-asset price. To approximate $I(x_0)$, we first truncate the integration region to an interval $[a, b]$ and then approximate the density f by using its Fourier cosine series expansion truncated to a finite number of terms

$$f_1(x|x_0) := \frac{A_0}{2} + \sum_{k=1}^{L} A_k(x_0) \cdot \cos(k\pi \frac{x-a}{b-a}) \tag{5.28}$$

with

$$A_k(x_0) := \frac{2}{b-a} \int_a^b f(x; x_0) \cos(k\pi \frac{x-a}{b-a}) dx, \quad k = 0, 1, \ldots. \tag{5.29}$$

The above cosine series coefficients can be approximated using the characteristic function $\phi(\cdot; x_o)$ of x as

$$A_k(x_0) \approx F_k(x_0) := \frac{2}{b-a} \Re \left\{ \phi(\frac{k\pi}{b-a}; x_0) \cdot \exp(-i \frac{ka\pi}{b-a}) \right\}, \tag{5.30}$$

where we have only an approximation not an equality in (5.30), which is due to the truncation of the integration region in the definition of the characteristic function. If we replace A_k with F_k in (5.28) and substitute the resulting sum into (5.27), we obtain the following approximation of $I(x_0)$ used in the COS method:

$$\hat{I}(x_0) := \frac{F_0}{2} + \frac{b-a}{2} \sum_{k=1}^{L} F_k(x_0) \cdot V_k, \tag{5.31}$$

where

$$V_k := \frac{2}{b-a} \int_a^b U(x) \cos(k\pi \frac{x-a}{b-a}) dx. \tag{5.32}$$

We can notice that $V_k,\ k = 1, 2, \ldots,$ are the cosine series coefficients of $U(x)$ on $[a, b]$, which suggests a close connection between the COS method and the PV method. Indeed, it is easy to verify that if we use the cosine series expansion of $U(x)$ on $[-l_x, l_x]$ in (S4–1) of the PV method , then (S4–2) produces a representation of the conditional expectation that has the same form as the expansion (5.31) when $a = -l_x$ and $b = l_x$.

Although the PV approach based on cosine functions and the COS method produce the same approximations, they have different starting points in expansions. This fact implies that the truncation error in both approaches can be controlled by approximating more closely either the value function or the density function. This result can also be derived from the common representation (5.31), where the rate at which the product $F_k(x_0) \cdot V_k$ decays to zero is faster than either $F_k(x_0)$ or V_k. However, the PV method is a broader framework than the COS method, since it offers alternative ways of approximating expectations.

To understand better how the PV approach can be used to improve the COS method, we first need to know some properties of the COS method when applied to the problem of pricing Bermudan options (Fang and Oosterlee (2008, 2009)):

(P1) Assume that the density of log-returns is smooth enough, typically we need only a small number of terms in its series expansion.

(P2) Assume that the series coefficients $V_k, k = 1, 2, ...,$ of the option values at the first early-exercise date are known, they can be calculated, for some options, very efficiently for other exercise dates through an induction formula combined with the FFT algorithm. In such cases, the computational complexity of the method is $O((M-1)L \log L)$, where M is the number of early-exercise dates and L is the number of terms in the series expansion of the density.

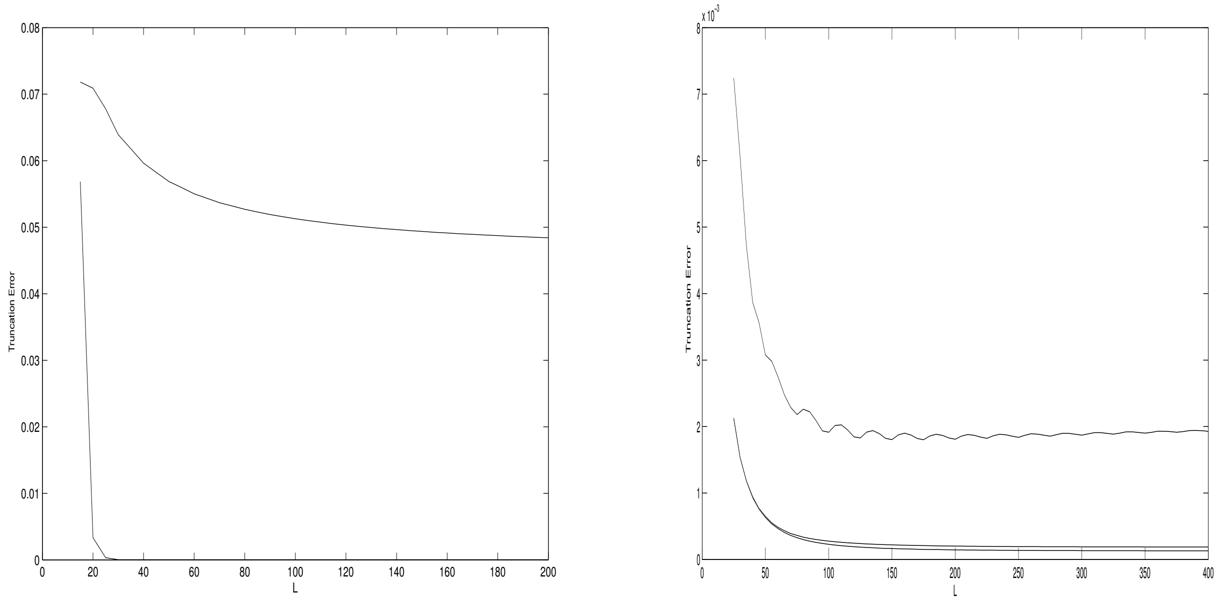
The first property (P1) depends on the smoothness of the density function of log-prices. In many financial applications, this density is often infinitely differentiable, and therefore its Fourier-cosine series expansion converges exponentially in the number of terms L on bounded intervals. As a result, in practice, we do not need to use large values of L to get accurate estimates of prices of European and Bermudan options. As we discuss below, however, for ratchet options the coefficients in Fourier-cosine series expansions show only algebraic convergence.

The second property (P2) of the COS method also leads to a significant reduction of its overall computational cost. To derive the induction formula, however, we need to know the form of the boundary that separates the continuation region from the exercise region at each exercise time. For some options and models, finding this boundary is equivalent to finding the point that separates the two regions (Fang and Oosterlee (2009)). However, the exact shape of the exercise region may be difficult to determine, especially for path-dependent options under a stochastic volatility model (for Heston model, see Ruijter and Oosterlee (2012)). Therefore, the feasibility of an induction formula and its effectiveness must be assessed on a case-by-case basis.

As mentioned earlier, for ratchet options, the exponential rate of convergence of the COS approach is typically not true. For these contracts, the density f corresponds to a truncated random variable, and it can be written in the form $f^t(x; x_0) := c_n f(x; x_0) 1_{[F,C]}(x)$, where 1_A is the indicator function of a set A, c_n is a normalizing factor, and the constants F and C with $F, C \in (a, b)$, $F < C$, determine the truncation levels. In this case, the density function has discontinuities inside the expansion region, and hence the truncation error in the cosine expansions of f^t will decay only algebraically.

Figure 5.1 shows some plots for the analysis of truncation errors (y-axis) with respect to the number of terms in expansions (x-axis). The left panel in Figure 5.1 depicts the truncation error, which is measured by the L^1-distance between functions, to demonstrate the difference between these two convergence rates. In this graph, the lower line represents the truncation error of the case where we approximate the standard normal density function on the interval $(-10, 10)$ using Fourier cosine expansions with L terms, while the upper line represents a similar error but for the normal density function truncated on the interval $(-2, 2)$. The difference is quite large, since for the former case we need only 16 terms to ensure that the error is less than 0.007, while for the latter case the distance is still larger than 0.01 even with 500 terms.

Figure 5.1: Truncation Error Analysis with respect to L



The PV method allows us to improve the convergence rates of the pricing methods based on characteristic functions by constructing better approximations of value functions. The justification for such a strategy is twofold. First, the value function has better smoothness properties than the density for ratchet options with surrender risk, since it is typically continuous and infinitely differentiable everywhere except for the points that separate the exercise and the continuation regions. Second, the PV method provides a more flexible framework for approximations than the COS method, since we are not constrained to cosine functions only. In particular, we can combine polynomials and trigonometric functions to construct more accurate approximations of the value function. Since these efficiency enhancing techniques usually depend on the problems at hand, below we only describe a general method that is applicable.

The truncation error when approximating $U(x)$ by the finite sum of trigonometric functions (5.15) is determined by the rate at which the coefficients $a_l,\ l = 0, 1, \ldots, L$, and $b_l,\ l = 1, \ldots, L$, defined in (5.16)–(5.17) converge to zero. Suppose that the function U is $(k-2)$-times continuously differentiable on $[-l_x, l_x]$ and its k-th derivative, $U^{(k)}$, is integrable on this interval. If, in addition, we also have

$$U(-l_X) = U(l_X),\quad U^{(1)}(-l_X) = U^{(1)}(l_X),\quad \ldots,\quad U^{(k-2)}(-l_X) = U^{(k-2)}(l_X), \tag{5.33}$$

then it can be shown that

$$|a_n| \leq \frac{F}{n^k} \quad \text{and} \quad |b_n| \leq \frac{F}{n^k}$$

for some sufficiently large constant F, which is independent of n. Thus, under the above assumptions, the algebraic index of converge is at least as large as k. A proof of this result based on simple integration by parts is presented in Boyd (1989) (Theorem 2.4). It is known that for American options the continuation value meets the payoff function smoothly, as long as the latter is smooth too (see, for example, Wilmott et al. 1993). Therefore, if we assume that the payoff function G is continuously differentiable for in-the-money region, then the first derivative of U will be continuous on $(-l_X, l_X)$, and hence the algebraic index of convergence will be at least 3 if the boundary conditions (5.33) hold for $k = 3$.

The value function $\hat{V}^e(t_{j+1}, \cdot, y)$ does not usually satisfy these conditions, but a simple remedy for this problem is to replace it with a function of the form:

$$\hat{V}^e_M(t_{j+1}, x, y) := \hat{V}^e(t_{j+1}, x, y) - p_{L_b}(x), \quad x \in [-l_X, l_X], \tag{5.34}$$

where the polynomial $p_{L_b}(x) := \sum_{l=0}^{L_b} d_l x^l$ is selected so that the resulting function $\hat{V}^e_M$ satisfies (5.33) for a prespecified value of k. A similar approach is proposed in Boyd (1989) for solving ordinary differential equations. In the context of the PV method, this technique is feasible since by (5.19) the integral of $p_{L_b}(x)$ can be expressed in terms of the characteristic function of f. In particular, we can ensure that the periodic extension of $\hat{V}^e(t_{j+1}, x, y)$ is continuous by using a linear function with

$$d_1 = \frac{\hat{V}^e(t_{j+1}, l_X, y) - \hat{V}^e(t_{j+1}, -l_X, y)}{2l_X}.$$

To illustrate the advantages of the proposed method over the cosine method, assume that U represents the log-prices of a Bermudan put option at time $t = 10$, where the option expires at $T = 11$ and its strike price is 1. We obtain U on the interval $(-5, 5)$ by using the Black-Scholes model and formula (5.10), where we assume that $\sigma = 30\%$ and $r = 5\%$. To ensure that the boundary conditions (5.33) are satisfied, we use (5.34) with either a linear function or a polynomial of order 3. Then we approximate U by using finite sums of trigonometric functions with $2L + 1$ terms for varying values of L. For comparison, we also approximate U using cosine expansions with L terms.

The right panel of Figure 5.1 shows the truncation errors when approximating a value function of a Bermudan option using either cosine expansions (top line), trigonometric expansions combined with a linear function (middle line), or trigonometric expansions combined with a polynomial of order three (bottom line). It shows that we can significantly improve the efficiency of the approximation based on trigonometric functions by combining it with a linear function. However, combining higher order polynomials with trigonometric functions does not improve the truncation error significantly over the case by combining a linear function. This can be explained by the fact that the continuation value does not have to meet the payoff function smoothly for Bermudan options, so the first derivative of U does not need to be continuous. We propose to combine trigonometric functions with polynomials of order one only when pricing options.

Regarding the second property (P2), the complexity of the COS method can be dramatically reduced for options where an efficient induction algorithm can be derived. In such cases the PV method is equally attractive, since it can be verified that the computational complexity remains unchanged if we combine cosine functions with a finite number of polynomials. Except for the cases discussed in the literature, it is unclear that whether the induction algorithm can be efficiently utilized or not for other options and/or models. For example, in Section 5.6.2 we demonstrate that for ratchet options under a regime switching model, the exercise region can be a union of two subsets. In the cases when the form of the exercise region prevents us from using an induction algorithm, the computational complexity of the COS method is only quadratic in the number of terms L. Therefore, in these situations we may consider any method of reducing L, such as the proposed PV approach.

5.4.2 The Least-Squares Method

Since in step (S4–2) of the PV method we can calculate conditional expectations by adding a finite number of explicit terms, the main challenge when using the method lies in step (S4–1), where we approximate the function $\hat{V}^e(t_{j+1}, \cdot, y)$ by a weighted sum of basis functions. If we estimate the coefficients α_l in (5.11) by the least-squares method, then this step becomes similar to the least-squares (LS) method proposed by Carrière (1996) and Longstaff and Schwartz (2001).

Similarly to the PV approach, the LS method is also based on the backward recursion (5.8)–(5.9). However, these two methods have several differences. First, we approximate the continuation value on the logarithmic scale in the PV method, while the scale is unchanged in the LS method. Second, instead of using the two steps (S4–1)–(S4–2) in the PV method, we approximate the continuation value (5.10) with the LS method in each time period by regressing $\hat{V}(t_{j+1}, S_{t_{j+1}})$ on functions of the form

$$\sum_{l=0}^{L} \beta_l m_l(S_{t_j}), \tag{5.35}$$

where $m_0, m_1, \ldots, m_L$ are given basis functions. For this, we need to sample a number of observations $(s^1_{t_j}, s^1_{t_{j+1}}), \ldots, (s^N_{t_j}, s^N_{t_{j+1}})$ from the joint distribution of $(S_{t_j}, S_{t_{j+1}})$ specified by the assumed dynamic of the price process [4]. Then we estimate the coefficients $\{\beta_0, \ldots, \beta_L\}$ by using the ordinary least-squares method. Approximating the continuation value in this way leads to two main drawbacks of the LS method:

(i) A proper selection of the basis functions and their number L is a challenging problem, since the function we are approximating, $s \to E[\hat{V}(t_{j+1}, S_{t_{j+1}})|S_{t_j} = s]$, is unknown. Therefore, it is difficult to find any diagnostic tool to asses the accuracy of our approximation at each time step and the accuracy of the resulting price of the option.

This point can be explained by the following arguments. Assume that the interest rate r is zero. By conditioning, we can derive the following equation

$$E\Big[\hat{V}(t_{j+1}, S_{t_{j+1}}) - \sum_{l=0}^{L} \beta_l m_l(S_{t_i})\Big]^2 \quad = \tag{5.36}$$

$$E[\hat{V}(t_{j+1}, S_{t_{j+1}}) - E\Big[\hat{V}(t_{j+1}, S_{t_{j+1}})|S_{t_j}]\Big]^2 + E\Big[\hat{C}(t_j, S_{t_j}) - \sum_{l=0}^{L} \beta_l m_l(S_{t_i})\Big]^2. \tag{5.37}$$

In the LS method, the least-squares procedure asymptotically converges to the minimization of (5.36) with respect to the parameters $\beta_0, \ldots, \beta_L$ as the number of simulated points N increases to infinity.

The above equation shows that minimizing (5.36) is equivalent to minimizing the second term in (5.37). This fact validates the use of regression techniques in the LS method, where we approximate the value of the option at the next time period $\hat{V}(t_{j+1}, S_{t_{j+1}})$ rather than approximating the continuation value $\hat{C}(t_j, S_{t_j})$, which is our objective of interest.

This decomposition also explains why it is difficult to assess the approximation error. It is known that the conditional expectation $E[\hat{V}(t_{j+1}, S_{t_{j+1}})|S_{t_j}]$ provides the best approximation of $\hat{V}(t_{j+1}, S_{t_{j+1}})$ in terms of the random variable S_{t_j} in the mean squared form. Typically, the difference between those two variables will be non-zero, which means the first term in (5.37) will be positive, regardless of the choice of the basis functions and the number of functions L we use. Thus, we cannot assess the goodness-of-fit and judge the accuracy of our approximation by the size of the residuals $\hat{V}(t_{j+1}, S_{t_{j+1}}) - \sum_{l=0}^{L} \beta_l m_j(S_{t_j})$.

[4] The required sample from the joint distribution of $(S_{t_j}, S_{t_{j+1}})$ can be simply obtained by simulating N paths of the price process, starting from the initial value S_0, and then sampling each path at t_j and t_{j+1}.

(ii) Since the explanatory variables in (5.35) are evaluated at randomly selected points drawn from the marginal distribution of S_{t_j}, we may not have enough observations in relevant regions for some options. Longstaff and Schwartz suggest that at each time, the regression should be carried out only in the region where the option is in the money. If the spot price S_0 belongs to the out-of-the-money region, many simulated paths of the process will remain in this region. Then, this method will be computationally inefficient.

(iii) Another potential drawback of the LS method when applied to models described in terms of characteristic functions of log-returns is due to the fact that direct sampling from such distributions is difficult and often requires numerical inversion procedures.

The PV method does not suffer from the drawbacks (i)–(iii) above. First, the function that we are approximating at each time period is known, at least over the mesh of points that we have selected. Therefore, it is possible to evaluate the distance between our approximation and the target function. Second, we can select points in any way that is suitable for the given function when we implement step (S4–1). For example, we propose to use uniformly spaced points, since this selection method allows us to utilize the FFT algorithm. Finally, the PV method does not require sampling from the distribution of log-returns.

However, the LS method is still attractive in the problems where the state space is high-dimensional, since in such cases methods based on transforms are still in their early stage of development.

5.5 Pricing Ratchet Options

The PV method described in Section 5.3 can be directly applied to price European and Bermudan options on the asset $\mathcal{S}$ defined by (5.1). In this section, we will show how to modify the algorithm to price simple ratchet EIAs with surrender risk.

For simple expositions, we assume that the contract has annual reset times or it credits the policyholder's account using annual returns of the underlying fund. The surrender option allows the policyholder to withdraw from the contract at the beginning of each year. Pricing a contract with surrender risk is the same as pricing a Bermudan option with pre-determined exercise times if we assume that the owner of the contract acts rationally and will surrender the contract only when it is optimal to do so. In addition, we assume that the contract payoff depends on a fund whose returns are described by model (5.1).

Suppose that $R_j := S_j/S_{j-1} = \exp(W_j)$, where W_j is defined in (5.1), represents the return during the annual period $[j-1, j]$, $j = 1, \ldots, N$, and the initial investment is \$1, then the payoff of a simple ratchet EIAs contract at the reset time j can be represented as:

$$U_j = 1 + \sum_{m=1}^{j} \max\Big\{F, \min\Big\{C, \alpha(R_m - 1)\Big\}\Big\}, \quad j = 1, \ldots, N, \tag{5.38}$$

with $U_0 = 1$, where F and C, $F < C$, are pre-specified local floor and cap levels and $\alpha > 0$ is a participation rate. The process (5.38) can also be represented in the form

$$U_j = U_{j-1} + R_j^*, \quad j = 1, \ldots, N, \tag{5.39}$$

with

$$R_j^* := \max\Big\{F, \min\Big\{C, \alpha(R_j - 1)\Big\}\Big\}. \tag{5.40}$$

Following Wei et al. (2013), we call R_j^* the ratchet interest rate in the time period $[j-1, j]$. For regime switching models, the variables, $R_1^*, R_2^*, \ldots,$ are not necessarily independent.

A more general form of the contract may consist of other features: a minimum garanteed value for owners, and a cancellation fee which is usually a pre-determined percentage of the annuity value. For example, Wei et al. (2013) choose the pre-determined percentage according to 2009 Annuity Fact Book as

$$\gamma(j) := 1 - \max\left\{\frac{N_c + 1 - j}{100}, 0\right\},$$

where N_c, $N_c < N$, is the number of years for which the cancellation fee is applied. If the contract is surrendered at the reset date j, $j = 1, \ldots, N-1$, its owner receives the amount

$$G(j, U_j) := \max\Big\{\gamma(j)U_j, \beta(1 + r_g)^j\Big\}, \tag{5.41}$$

where $r_g \geq 0$ is the minimum guaranteed annual interest rate and β is a percentage of the initial premium. If the contract is not surrendered earlier, then the owner receives the amount $G(N, U_N)$ at maturity.

The value of the above contract can be represented as

$$V := \max_{\tau \in \mathcal{T}} E\Big[e^{-\tau r} G(\tau, U_\tau)\Big],$$

where τ is a stopping time taking values in the set $\{1, \dots, N\}$.

To obtain the price of the contract under a regime switching model, we use the PV method with a specialized backward recursive procedure of (5.9)- (5.10) and modified steps of (S4–1)–(S4–2). First, replace the process $\mathcal{S}$ with the sequence $U_1, \dots, U_N$, which is a direct consequence of the form of the payoff (5.41). To add the regime process $\mathcal{Y}$, consider at the reset time $N-1$, the value of the contract $V_R(N-1, u, y)$ is equal to the maximum of $G(N-1, u)$ with $u = U_{N-1}$, and the continuation value is

$$C_R(N-1, u, y) := e^{-r} E\Big[\max\Big\{\gamma(N)[U_{N-1} + R_N^*], \beta(1+r_g)^N\Big\} | U_{N-1} = u, Y_{N-1} = y\Big]. \tag{5.42}$$

We have three remarks for the representation (5.42) with model (5.1).

Remark 5.5.1.

(i) When U_{N-1} and Y_{N-1} are known, then the ratchet interest rate R_N^* is independent of the path of the process (5.38), and is determined only by the distribution of R_N based on (5.40).

(ii) The variables U_{N-1} and R_N^* in (5.42) are in an additive way, so we only need to approximate $V_R(N-1, u, y)$ on the original not logarithmic scale when we apply (S4–1) of the PV method. In addition, $V_R(N-1, u, y)$ only depends on the payoff at time N and does not depend on the future regime Y_N, this implying that we need only one approximation at this step.

(iii) To apply step (S4–2) of the method, we need to determine the characteristic function of R_N^* in each regime y.

Now consider the value of the contract $V_R(j, u, y)$ at time j, $j < N-1$. Then, the continuation value (5.42) generalizes to

$$C_R(j, u, y) := e^{-r} E\Big[V_R(j+1, U_j + R_{j+1}^*, Y_{j+1}) | U_j = u, Y_j = y\Big]. \tag{5.43}$$

Equation (5.43) has similar formulation as (5.10). Thus, we can apply the PV method to price ratchet EIAs with surrender risk only if we can determine the characteristic function of the variable R_{j+1}^* conditioned on the current regime. To summarize steps (S4–1)–(S4–2) of the PV method that are specialized to a ratchet option, let us define $\hat{V}_R(j+1, \cdot, \cdot)$ as an approximation of $V_R(j+1,\cdot,\cdot)$ obtained through the backward recursion, with $\hat{V}_R(N,u,y) \equiv V_R(N,u,y) = G(N,u)$. Then, at time t_j and for regime $k \in \mathbf{S}$, the continuation value is approximated by:

$$\begin{aligned}
E\Big[\hat{V}_{R,\mathcal{M}_{(l_X,L)}}(j+1, U_j + R^*_{j+1}, Y_{j+1})|U_j = u, Y_j = k\Big] &= \frac{1}{2}E[a_0(Y_{j+1})|Y_j = k] \\
+\sum_{l=1}^{L}\Big\{\big[\bar{a}_l(k)R_l(k) + \bar{b}_l(k)I_l(k)\big]\cos(\frac{\pi l}{l_X}u) + \big[\bar{b}_l(k)R_l(k) - \bar{a}_l(k)I_l(k)\big]\sin(\frac{\pi l}{l_X}u)\Big\}&, \quad (5.44)
\end{aligned}$$

where $\bar{a}_l(k)$, $\bar{b}_l(k)$ are defined in (5.23) and

$$a_l(y) := \frac{1}{l_X}\int_{-l_X}^{l_X} \hat{V}_R(j+1,x,y)\cos(\frac{\pi l x}{l_X})dx, \quad l = 0,1,2,\ldots,L, \qquad (5.45)$$

$$b_l(y) := \frac{1}{l_X}\int_{-l_X}^{l_X} \hat{V}_R(j+1,x,y)\sin(\frac{\pi l x}{l_X})dx, \quad l = 1,2,\ldots,L, \qquad (5.46)$$

$$R_l(k) := \Re\Big(\Psi_{R,k}(\frac{\pi l}{l_X})\Big) \quad \text{and} \quad I_l(k) = \Im\Big(\Psi_{R,k}(\frac{\pi l}{l_X})\Big), \quad l = 1,2,\ldots,L. \qquad (5.47)$$

To use the above formulas, we need the characteristic function $\Psi_{R,k}(\cdot)$ of the ratchet interest rate R^*_{j+1} in each regime k. Since the method that we propose to find $\Psi_{R,k}(\cdot)$ has the same form in each regime, we explain how to obtain the characteristic function Ψ_{R^*} of a variable R^* defined by

$$R^* := \max\Big\{F, \min\Big\{C, \alpha[R-1]\Big\}\Big\} = \max\Big\{F, \min\Big\{C, \alpha[e^W - 1]\Big\}\Big\},$$

where $R = \exp(W)$ and the characteristic function of W is known as Ψ_W.

Let us denote the density function of W by $f_W(\cdot)$. Using (5.28)–(5.30), we can approximate $f_W(\cdot)$ over $[a,b]$ by its Fourier cosine series expansion truncated to a finite number of terms

$$f_W(x) \approx \sum_{k=0}^{L^*}{}' F_k \cos(k\pi\frac{x-a}{b-a}), \qquad (5.48)$$

where F_k is given by (5.30) and $\sum'$ means that the first coefficient in the sum is divided by two. Then,

$$\begin{aligned}
\Psi_{R^*}(z) &= \int_{-\infty}^{\infty} e^{iz\max(F,\min(C,\alpha[e^w-1])} f_W(w)dw \\
&\approx \int_{-\infty}^{\infty} e^{iz\max(F,\min(C,\alpha[e^w-1])} \sum_{k=0}^{L^*}{}' F_k \cos(k\pi\frac{w-a}{b-a})dw \\
&\approx \sum_{k=0}^{L^*}{}' F_k \Bigg(e^{izF} \int_{\tilde{a}}^{\ln(F/\alpha+1)} \cos(k\pi\frac{x-a}{b-a})dx + \int_{\ln(F/\alpha+1)}^{\ln(C/\alpha+1)} e^{iz\alpha(e^x-1)} \cos(k\pi\frac{x-a}{b-a})dx \\
&\qquad + e^{izC} \int_{\ln(C/\alpha+1)}^{\tilde{b}} \cos(k\pi\frac{x-a}{b-a})dx \Bigg), \qquad (5.49)
\end{aligned}$$

where $[\tilde{a},\tilde{b}]$ defines a bounded integration region, which we can take the same as $[a,b]$. Let $q_1 = \ln(F/\alpha+1)$ and $q_2 = \ln(C/\alpha+1)$. Using the trapezoidal rule, we approximate the middle integral in (5.49) by

$$\begin{aligned}
&\frac{h_{q_1,q_2}}{2}\Big(e^{izF}\cos(k\pi\frac{q_1-a}{b-a}) + e^{izC}\cos(k\pi\frac{q_2-a}{b-a})\Big) \\
&\qquad + h_{q_1,q_2}\sum_{m=1}^{n-1}\Big(e^{iz\alpha(e^{q_1+mh_{q_1,q_2}}-1)}\cos(k\pi\frac{q_1+mh_{q_1,q_2}-a}{b-a})\Big), \qquad (5.50)
\end{aligned}$$

where n denotes the number of equally spaced integration nodes and $h_{q_1,q_2} := (q_2-q_1)/n$ is the distance between two consecutive nodes. By combining (5.49) and (5.50), we approximate the real and imaginary parts of $\Psi_{R^*}(z)$ respectively by

$$\begin{aligned}
&\frac{F_0}{2}\Bigg[\cos(zF)(q_1-a) + \cos(zC)(b-q_2) + \frac{h_{q_1,q_2}}{2}\Big(\cos(zF)+\cos(zC)\Big) \\
&+ h_{q_1,q_2}\sum_{m=1}^{n-1}\cos\Big(z\alpha(e^{q_1+mh_{q_1,q_2}}-1)\Big)\Bigg] \\
&+ \sum_{k=1}^{L^*} F_k\Bigg[\cos(zF)\sin(k\pi\frac{q_1-a}{b-a})\frac{b-a}{k\pi} - \cos(zC)\sin(k\pi\frac{q2-a}{b-a})\frac{b-a}{k\pi} \\
&+ \frac{h_{q_1,q_2}}{2}\Big(\cos(zF)\cos(k\pi\frac{q_1-a}{b-a}) + \cos(zC)\cos(k\pi\frac{q_2-a}{b-a})\Big) \\
&+ h_{q_1,q_2}\sum_{m=1}^{n-1}\Bigg(\cos\Big(z\alpha(e^{q_1+mh_{q_1,q_2}}-1)\Big)\cos\Big(k\pi\frac{q_1+mh_{q_1,q_2}-a}{b-a}\Big)\Bigg)\Bigg] \qquad (5.51)
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{F_0}{2}\Bigg[\sin(zF)(q_1-a)+\sin(zC)(b-q_2)+\frac{h_{q_1,q_2}}{2}\Big(\sin(zF)+\sin(zC)\Big)\\
&+h_{q_1,q_2}\sum_{m=1}^{n-1}\sin\Big(z\alpha(e^{q_1+mh_{q_1,q_2}}-1)\Big)\Bigg]\\
&+\sum_{k=1}^{L^*}F_k\Bigg[\sin(zF)\sin(k\pi\frac{q_1-a}{b-a})\frac{b-a}{k\pi}-\sin(zC)\sin(k\pi\frac{q2-a}{b-a})\frac{b-a}{k\pi}\\
&+\frac{h_{q_1,q_2}}{2}\Big(\sin(zF)\cos(k\pi\frac{q_1-a}{b-a})+\sin(zC)\cos(k\pi\frac{q_2-a}{b-a})\Big)\\
&+h_{q_1,q_2}\sum_{m=1}^{n-1}\Bigg(\sin\Big(z\alpha(e^{q_1+mh_{q_1,q_2}}-1)\Big)\cos\Big(k\pi\frac{q_1+mh_{q_1,q_2}-a}{b-a}\Big)\Bigg)\Bigg].
\end{aligned}
\tag{5.52}
$$

5.6 Numerical Examples

In this section, we illustrate the PV pricing method using different models and options. In Section 5.6.1, we price European and Bermudan options under models where we assume the Markov chain $\mathcal{Y}$ stays in one regime only, and we compare our prices with the results available in the literature based on alternative methods. In Section 5.6.2, we use three two-regime switching models to price ratchet EIAs.

5.6.1 European and Bermudan Options under Constant Regime

We first assume that the model parameters are constant in time, that is, the Markov process $\mathcal{Y}$ stays in one regime. Under this assumption we use the PV approach to price European and Bermudan options on an underlying asset that follows (5.1) under six different selections of the Lévy process $\mathcal{L}$. In addition to the Black-Scholes model (BS), the Merton model and the variance gamma model (VG) that we discuss in Section 1.1.2, we also consider the jump-diffusion model proposed by Kou (2002), the normal inverse Gaussian process (NIG) and the tempered α-stable process (TS) (for more information about these processes, see, for example, Tankov (2003)). By including the NIG model, we can compare our prices with the results obtained by Këllezi and Webber (2004), where the authors propose a lattice method to price Bermudan options on an underlying asset whose log-prices follow a Lévy process. In the Appendix 5.A, we present the characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ for all six Lévy processes.

The model parameters that we use are listed below, where the parameters for the NIG and VG models are the same as those used in Këllezi and Webber (2004). For other models, we have obtained the parameters by matching the first five central moments of one year log-returns with those from the VG model.

Black-Scholes model:

$$\sigma = 0.131.$$

Merton model:

$$\sigma = 0.067, \quad \lambda = 1.618, \quad \sigma_J = 0.032, \quad \mu_J = -0.086.$$

Kou model:

$$\sigma = 0.065, \quad p = 0.090, \quad \lambda = 4.136, \quad \lambda_+ = \lambda_- = 24.221.$$

VG model:

$$\nu = 0.2, \quad \sigma = 0.12, \quad \theta = -0.14.$$

NIG model:

$$\alpha = 28.421, \quad \beta = -15.086, \quad \delta = 0.317 \quad \mu_0 = 0.059.$$

TS model

$$\alpha = 0.273, \quad c_+ = 2.093, \quad c_- = 1.952, \quad \lambda_+ = 38.209, \quad \lambda_- = 16.05.$$

We also assume that the continuously compounded interest rate $r = 10\%$ and the spot price or S_0 of the underlying security is 100.

Table 5.1 shows the prices of European calls and puts obtained using the PV method. The options have one-year time to maturity, $T = 1$, and varying strike prices. Other parameters we use including the truncation parameter $l_X = 4$ and the number of basis functions $L = 200$, which corresponds to 401 ($= 2 * L + 1$) basis functions in the set (5.14). At each time period, the continuation value is approximated by equation (5.15), where the coefficients are obtained by applying a numerical integration method to the integrals in (5.16) and (5.17) with $M = 1000$ nodes. This continuation value can be seen as the price of the option at this period. For a European option, its price is the discounted expectation of the payoff at maturity, and it can be obtained in one step only; however, we use the recursive algorithm to capture the total error accumulated in all time steps. From Table 5.1, we can see that the prices of the NIG and VG models are consistent with those presented by Këllezi and Webber (2004) up to 3 decimal places.

Table 5.1: Prices of European Call and Put options with Maturity One Year

Option	Strike	Model					
		BS	Merton	Kou	NIG	VG	TS
Call	90	18.855	19.100	19.103	19.093	19.099	19.098
	100	11.097	11.413	11.374	11.360	11.370	11.366
	110	5.428	5.496	5.425	5.437	5.430	5.430
Put	90	0.291	0.536	0.538	0.529	0.535	0.533
	100	1.581	1.896	1.858	1.844	1.854	1.849
	110	4.960	5.029	4.957	4.969	4.962	4.962

Table 5.2 shows the prices of at-the-money calls and puts with different values of l_X and L in order to see the impacts of the selection of l_X and L on the accuracy of the method, where we only consider the TS and Kou models. For a wide range of parameter values, the prices are reasonably close, which suggests the PV method is stable. The results also show that pricing options with unbounded payoff functions may be more difficult. Take a call option as an example, the difference between end-values of the payoff function over $[-l_X, l_X]$ becomes larger as the truncation parameter l_X increases, which results in larger discontinuity of the periodic extension of the payoff function. As a consequence, for larger values of l_X, we need to increase the number of basis functions to guarantee the convergence of the prices. We have repeated this experiment for other specifications of our models, and the results are very similar.

Table 5.2: Prices of European at-the-money Calls and Puts for Different Selections of the Truncation Parameter l_X and the Number of Basis Functions (equal to $2 * L + 1$)

Option	L	TS Model			Kou Model		
		$l_X = 1$	$l_X = 4$	$l_X = 10$	$l_X = 1$	$l_X = 4$	$l_X = 10$
Call	200	11.366	11.366	-6.306	11.374	11.374	-8.329
	400	11.366	11.366	11.383	11.374	11.374	11.382
	600	11.366	11.366	11.366	11.374	11.374	11.374
Put	200	1.849	1.849	1.851	1.858	1.858	1.860
	400	1.849	1.849	1.849	1.858	1.858	1.858
	600	1.849	1.849	1.849	1.858	1.858	1.858

Table 5.3 shows the prices of Bermudan put options, where the option could be exercised at one of ten equally spaced time periods. This assumption is consistent with the setup used by Këllezi and Webber (2004). Other parameters we use including $L = 600$, $M = 1000$ and $l_X = 4$. We find that for all the models except for the TS, the prices converge to their limits quite fast with respect to the number of basis functions. In particular, we need 401 basis functions to guarantee that the results are within 0.01% of their limits for these five models. For the TS model, the error is 0.03% for 401 basis functions, and it becomes less than 0.003% for 601 basis functions. We have compared the results for the NIG and VG models with those obtained by Këllezi and Webber (2004), and they are consistent within 0.02%. The results also suggest that while the models produce similar prices, the BS model can significantly underprice deep out-of-the-money options.

Table 5.3: Prices of Bermudan Put Options with Maturity One Year and 10 Equally Spaced Exercise Opportunities

Strike price		BS	Merton	Kou	NIG	VG	TS
	90	0.376	0.766	0.768	0.745	0.761	0.756
	95	1.031	1.571	1.539	1.496	1.526	1.515
	100	2.443	3.008	2.889	2.844	2.882	2.866
K	105	5.097	5.292	5.158	5.173	5.170	5.166
	110	9.116	9.004	9.019	9.0340	9.041	9.039
	115	13.875	13.856	13.877	13.865	13.876	13.873
	120	18.807	18.806	18.814	18.807	18.810	18.809

5.6.2 Ratchet EIAs under Regime Switching Models

Here we present the implementation results of the PV method to price ratchet EIAs. We assume that the underlying fund follows model (1.1) under the three specification of the distribution of $Z(\underline{\xi})$ that are formulated in Section 1.1.2. We shall refer to them respectively as the regime switching Black-Scholes (RSBS), Merton (RSM), and variance gamma (RSVG) models.

For parameters of the RSVG model, we use one set of values presented in Konikov and Madan (2002), where the authors fit continuous-time two-state VG Markov models to time series data on daily returns of 22 individual stocks and indices. They are:

$$\nu_1 = 0.002, \quad \nu_2 = 0.002 \quad \sigma_1 = 0.780, \quad \sigma_2 = 0.302, \quad \theta_1 = 2.442, \quad \theta_2 = -1.106,$$

and

$$p_{11} = 0.434, \quad p_{21} = 0.434.$$

The parameters of the RSBS and RSM models are obtained by matching prices of European put options with those obtained from the RSVG model. For the RSBS model, we use options with one-year maturity and strikes equal to 90, 100, 110, and 120, while for the RSM model, the strikes are 85, 90, 95, 100, 105, 110, 115, and 120. By using the Euclidean distance, we matched the prices at the level of error less than 0.01%. The resulting parameters are:

RSBS[5]:

$$\sigma_1 = 0.548, \quad \sigma_2 = 0.654, \quad p_{11} = 0.482, \quad p_{21} = 0.743,$$

RSM[6]:

$$\mu_J = -0.255, \quad \sigma_1 = 0.790, \quad \sigma_2 = 0.391, \quad \lambda_1 = 0.004, \quad \lambda_2 = 0.213,$$

$$\sigma_J = 0.677, \quad p_{11} = 0.441, \quad p_{21} = 0.238.$$

Table 5.4 shows the prices of ratchet options with surrender (S) risk and without surrender (NS) risk for different values of F and C. We assume $r = 5\%$, $T = 10$, $\alpha = 1$, $r_g = 3\%$, $\beta = 1$, and $N_c = 8$. The prices are obtained by applying the PV method with the basis functions given by the set $\mathcal{M}_{(l_X,L)}$ with $l_X = 2$ and $L = 400$. To calculate the coefficients (5.16)–(5.17), we use the trapezoidal quadrature rule with $M = 5000$ nodes. To compute the real and imaginary parts of the characteristic function of R^* in equations (5.51) and (5.52), we use $L^* = 200, a = -5, b = 5$, and $n = 200$. In the last row, we present prices for large absolute values of C, which can be seen as an approximation of the case

[5]Note that $\sigma_1 < \sigma_2$ in the RSBS model, but $\sigma_1 > \sigma_2$ in the RSVG and RSM models. This may be explained by the fact that we obtain the model parameters by matching the option prices, and the transition probabilities p_{11} and p_{21} are quite different for those models, and hence the moments could be similar though the standard deviations, σ_1 and σ_2, are quite different for those models.

[6]We assume that μ_J and σ_J stay in only one regime (remain constants) instead of two as considered in Chapters 2-4 for the RSM model. Because it is more convenient for us to obtain 10 instead of 12 parameters. In addition, our goal in this chapter is to price options, but not to obtain the most flexible model.

without any truncation in the right tail. To preserve the accuracy of the pricing method, we need to increase l_X to 8 in this case .

For comparison purposes, we also provide prices of ratchet options without surrender risk obtained from the Monte Carlo (MC) simulation with 100,000 sampled paths in the RSBS model. The numbers in the parentheses are the standard deviations of the MC estimates for the prices. Overall, the results agree with those obtained from the PV method, except for the case $C = 5$ where the error is about 2%. This pattern can be explained by the fact that for a larger C, it is more difficult to obtain accurate approximations of the characteristic function of the ratchet interest rate R_j^*.

Table 5.4: Prices of Ratchet Options Under Different Regime Switching Models

Floor	Cap	RSBS		RSM		RSVG		RSBS–MC
		NS	S	NS	S	NS	S	NS
0%	10%	0.866	0.980	0.866	0.980	0.869	0.980	0.866(0.000)
0%	20%	1.047	1.061	1.043	1.057	1.045	1.059	1.047(0.000)
0%	30%	1.212	1.218	1.202	1.208	1.196	1.202	1.211(0.000)
0%	500%	2.155	2.197	2.113	2.191	2.008	2.080	2.199(0.003)

Table 5.5 shows the prices for different values of l_X, which is in order to see the impact of the parameter on the accuracy of the pricing procedure. The results become stable when $l_X = 2$ except the cases when C is large, in which l_X needs to be at least 8. Thus, the presence of the floor and the cap requires more terms in Fourier expansions of the payoff function, but it also reduces the integration region, which makes some steps in the algorithm more accurate.

Table 5.5: Prices of Ratchet Options with Surrender Risk for Different l_X

Floor	Cap	RSBS			RSM			RSVG		
		$l_X = 2$	$l_X = 6$	$l_X = 10$	$l_X = 2$	$l_X = 6$	$l_X = 10$	$l_X = 2$	$l_X = 6$	$l_X = 10$
0%	10%	0.980	0.976	0.981	0.980	0.977	0.980	0.980	0.977	0.980
0%	20%	1.061	1.059	1.059	1.057	1.055	1.056	1.059	1.058	1.059
0%	30%	1.217	1.219	1.220	1.208	1.209	1.210	1.202	1.203	1.204
0%	500%	1.520	2.153	2.201	1.495	2.125	2.201	1.463	2.023	2.088

In Figure 5.2, we compare the rates of convergence of the PV method based on two different basis functions. One is only based on trigonometric functions (continuous line), and the other one is based on trigonometric functions with a linear function (dotted line), which is described in Section 5.4.1. For this comparison, we use an RSBS model to price a ratchet option with surrender risk, where $F = 0,$ and $C = 0.1$. As the graph suggests, the addition of the linear function allows us to reduce L from 400 to 100. Although the improvement is not dramatic, it is quite noticeable given that the method is easy to implement and does not require any additional computational time.

Figure 5.2: Prices of a Ratchet Option with Surrender Risk under the RSBS Model

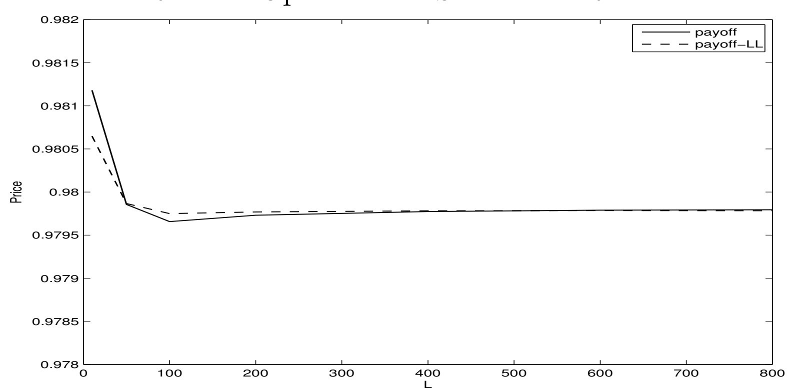
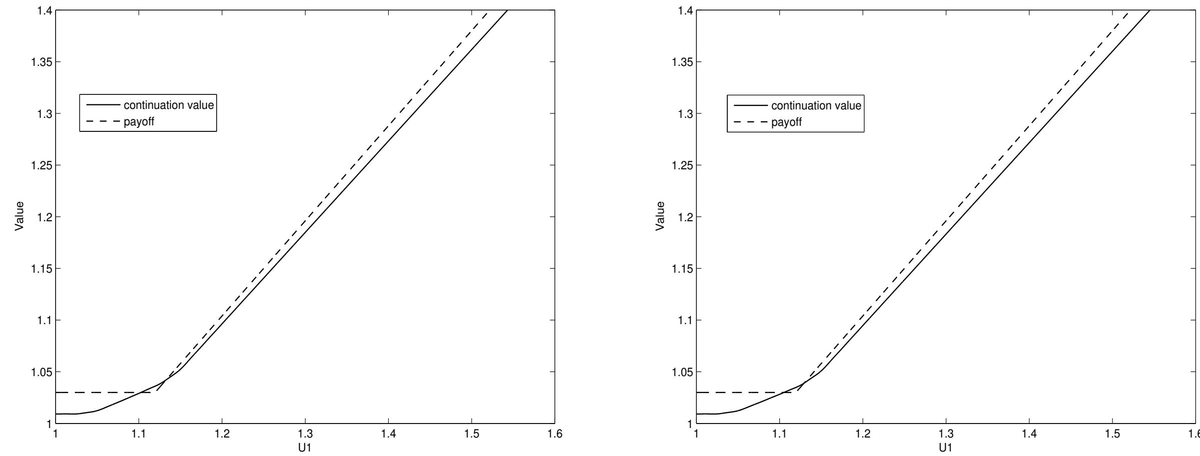


Figure 5.3 shows the continuation values (continuous lines) and the payoff functions (dotted lines) at time $t = 1$ for a ratchet option with surrender risk where $F = 0$ and $C = 0.1$ under an RSBS model. The plot shows that in each regime the region where the option should be surrendered is a union of two subregions. This fact has direct consequences on possible improvements of the COS and the PV pricing methods, which is left for future research.

Figure 5.3: Exercise Regions at time $t = 1$ for a Ratchet Option with Surrender Risk under the RSBS Model in Regime 1 (left panel) and Regime 2 (right panel)



5.7 Conclusions of Chapter 5

In Chapter 5, we propose a pricing method called the PV method to price Bermudan options. The method is based on the dynamic programming approach, where we calculate the conditional expectation by representing the current value of the option with Fourier series expansions and then applying the characteristic function of the underlying process. To improve the rate of convergence of option prices, we describe a smooth trick where we combine trigonometric functions with polynomial functions to approximate value functions.

To understand the advantages of the PV method, we compare it with the COS method proposed by Fang and Oosterlee (2008) and the LS method proposed by Carrière (1996) and Longstaff and Schwartz (2001). The PV method can recover the COS method by using cosine functions as its basis functions, and it performs more flexibly than the COS method especially for the options whose density functions are not smoother than their payoff functions.

Moreover, we apply the PV method to European and Bermudan options under constant regime models in Lévy processes. The resulting prices are consistent with our benchmarks. Then we use the PV method to price ratchet EIAs with and without surrender risk under two-regime switching models, where the smooth trick is proven to be useful.

Finally, we find that ratchet EIAs with surrender risk may have two exercise regions. This observation has direct consequences on possible improvements of the PV and COS pricing methods.

Chapter 6

Sampling Conditioned Processes for a Regime Switching Black-Scholes Model

In this chapter, we propose an algorithm to sample conditioned processes from a regime switching Black-Scholes model. The sampling method that we propose can be applied to different problems, including filtering of a hidden Markov process and characterization of an optimal static hedging option. In the second part of this chapter, we describe the latter in detail.

The main computational challenge in finding the optimal hedging option in a recently proposed method by Kolkiewicz (2016) is the generation of paths from conditioned processes, called bridges or pinned down processes. Specifically, let $\{S_t, t \in [0,T]\}$ be a stochastic process representing prices of a traded security. Suppose we know that $S_s = x$ and $S_T = z$, where $s \in [0,T)$. We want to generate a value of the process at time t, where $t \in (s,T)$. Once we have the value, we repeat the procedure. This time we assume that S_t and S_T are known and the task is to generate a value at time u, where $u \in (t,T)$. We continue this procedure until we have an enough values along this path. In real applications we may need to generate 250 values to get a skeleton of a single path of the conditioned process. This will be enough to find, for example, an average along each path. As we discuss later, to find an optimal hedging option we have to repeat this simulation for a large number of terminal values z. This procedure is quite straightforward for a geometric Brownian motion, as it reduces to sampling Brownian bridges (see Kolkiewicz (2016) for more details). However, sampling bridges for processes that include a hidden Markov chain is much more difficult.

In this chapter, we assume that the stochastic process $\{S_t\}$ follows a Black-Scholes model with volatility that follows a continuous-time Markov chain. Our objective is to sample the process $\{S_t\}$ given its initial and terminal values and the initial state of volatility. Rigorously speaking, this is not a bridge sampling problem as we do not assume that we know the terminal value of the Markov chain. Therefore, we call it the problem of sampling a conditioned process. Literature on methods of bridge sampling for such models is still scarce. Bridge sampling method was introduced to statistics by Meng and Wong (1996) to solve the problem of computing the ratio of the normalizing constants that can be used in likelihood and Bayesian inference, which was further discussed by Gelman and Meng (1998). However, the method is based on i.i.d. samples, and hence cannot be applied to our problem. Hobolth and Stone (2009) propose an algorithm to sample states from a continuous-time Markov chain given the initial and terminal states. However, this is a different problem from ours, because in our problem we only know S_T but not the terminal value of the Markov process[1]. Delyon and Hu (2006) propose some algorithms to sample paths from the distribution of a diffusion process given the terminal value, but their methods are only applicable to some particular diffusion processes.

This chapter is organized as follows. In Section 6.1, we present an algorithm to sample conditioned processes from a two-regime switching Black-Scholes model, which is called the conditioned process sampling algorithm. In Sections 6.2 and 6.3, we apply the proposed sampling algorithm to the problems of pricing and static hedging of Asian options respectively. In Section 6.4, we briefly discuss the application of the proposed conditioned process sampling algorithm to the problem of filtering of a hidden Markov process. Section 6.5 concludes.

6.1 Sampling Method for the Conditioned Process

In this section, we propose an algorithm to sample conditioned processes from a regime switching Black-Scholes model where the volatility follows a continuous-time Markov chain.

Here we assume that under a risk-neutral measure Q the price process $\{S_t\}$ of the underlying asset follows the Black-Scholes model:

$$dS_t = rS_t dt + v_t S_t dW_t, t \in [0, T], \tag{6.1}$$

[1]As suggested by my committee member, Professor Tony Wirjanto, it is often only possible to observe the log-returns of the underlying asset, i.e., the integrated HMM at discrete time points, for example, some asset prices are only quoted on a daily basis.

where r is a risk-free continuously compounded interest rate, the volatility $\{v_t\}$ is independent of the standard Brownian motion $\{W_t\}$ and follows a continuous-time Markov chain. conditional on the volatility path, we can obtain the following result by Itô's lemma:

$$\ln \frac{S_t}{S_0} \sim N(rt - \frac{1}{2}\int_0^t v_s^2 ds, \int_0^t v_s^2 ds), \quad t \in [0, T]. \tag{6.2}$$

This implies that for any $u < t$, $u, t \in [0, T]$, we have

$$\ln \frac{S_t}{S_u} \sim N(r(t-u) + \int_u^t \frac{1}{2} v_s^2 ds, \int_u^t v_s^2 ds). \tag{6.3}$$

We assume that there are two states (i.e., σ_1 and σ_2, $\sigma_1 \neq \sigma_2$ and $\sigma_1, \sigma_2 > 0$), and the infinitesimal generator of the Markov chain can be written as

$$A = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}, \tag{6.4}$$

where $\lambda_1, \lambda_2 > 0$. For convenience, we also assume that the initial state is 1 (i.e., $v_0 = \sigma_1$) throughout this chapter. Indeed, in practice the initial state can be estimated using a calibration procedure.

Our goal is to generate a path of $\{S_{t_i}, i = 1, ..., n\}$ given the initial value S_0 and the terminal value S_T, where $\{t_i, i = 1, ..., n\}$ is a set of pre-determined discrete times such that $0 < t_1 < ... < t_n = T$. Before we present our proposed sampling method, we have to introduce some additional notations. Define

$$U_{t_1,t_2} := \int_{t_1}^{t_2} v_s^2 ds, \quad 0 \leq t_1 < t_2 \leq T, \tag{6.5}$$

and we denote $U_{0,T}$ by U as

$$U \equiv U_{0,T} = \int_0^T v_s^2 ds. \tag{6.6}$$

In addition, let us denote the total time spent in regimes 1 and 2 over the time interval $[0, T]$ by T_1 and T_2 respectively. Note that $T_1 + T_2 = T$.

We first find the characteristic function of the joint distribution of (T_1, T_2), $\phi_{T_1,T_2}(z_1, z_2) := E[e^{iz_1T_1 + iz_2T_2}]$, where $z_1, z_2 > 0$. Let π_0 denote the row vector of the initial (at time 0) probabilities for the chain, and $\mathbf{1}$ denote the 2-dimensional unit column vector. Let D be a diagonal matrix with $z := (z_1, z_2)$ being its diagonal entries. Following the results in Elliott et al. (2005), we have

$$\phi_{T_1,T_2}(z_1, z_2) = E[e^{iz_1T_1+iz_2T_2}] = \pi_0 e^{(A+iD)T}\mathbf{1} = e^{\lambda T} + e^{(iz_1-\lambda)T}. \tag{6.7}$$

If $z_2 = 0$, then we have the characteristic function of T_1 which we denote by ϕ_{T_1}. Indeed, ϕ_{T_1} has the same form as (6.7). This result can be also used to obtain the characteristic function of U, which is needed in the sampling algorithm proposed later. Based on our model assumptions, (6.6) can be represented as

$$U = \sigma_1^2 T_1 + \sigma_2^2 T_2 = \sigma_1^2 T_1 + \sigma_2^2(T - T_1) = \sigma_1^2(T - T_2) + \sigma_2^2 T_2. \tag{6.8}$$

Therefore, the characteristic function of U can be written as

$$\phi_U(z) = \phi_{T_1,T_2}(\sigma_1^2 z, \sigma_2^2 z) = e^{\lambda T} + e^{(i\sigma_1^2 z-\lambda)T},\ z > 0. \tag{6.9}$$

By the definition of characteristic functions, if $z = 0$, then $\phi_U(z) = 1$.

We first assume that $\lambda_1 = \lambda_2 = \lambda$. In this case, the total number of changes (jump events) in regime can be seen as following a Poisson process with the arrival intensity λ. Denote the arrival rate of the jump events during the time interval $[0, \Delta]$ by $\lambda_\Delta := \lambda \cdot \Delta$, where $\Delta \in \mathbb{R}^+$. Define N_Δ as the total number of events that occur during the time interval $[0, \Delta]$. Then N_Δ has a Poisson distribution with rate λ_Δ, that is $N_\Delta \sim POI(\lambda_\Delta)$. Define $0 = X_0 < X_1 < ... < X_{N_\Delta} < X_{N_\Delta+1} = T$ as the corresponding event times.

In Algorithm 6.1 below, we outline the main steps of the algorithm we propose to generate a path of $\{S_{t_i}, i = 1, ..., n\}$ given $S_0 = s_0$ and $S_T = s_T$ when $\lambda_1 = \lambda_2 = \lambda$. Figure 6.1 shows the flowchart of Algorithm 6.1 that contains four main steps. For each step, we derive the conditional distribution that we need to sample from by Bayes' theorem, and we provide details in Sections 6.1.1-6.1.4. In Section 6.1.5, we extend our algorithm to the case when $\lambda_1 \neq \lambda_2$.

Figure 6.1: Flowchart of Algorithm 6.1

$$\boxed{U|S_T} \longrightarrow \boxed{N_T|U} \longrightarrow \boxed{X_1, ..., X_{N_T}|U, N_T} \longrightarrow \boxed{S_{t_i}|S_0, S_T; v_{t_i}}$$

Algorithm 6.1

Require: Given $S_0 = s_0$, $S_T = s_T$, $v_0 = \sigma_1$, $\Delta = T$ and the infinitesimal generator of the Markov chain.

Output: A path of $\{S_{t_i}, i = 1, ..., n\}$ under the assumption that $\lambda_1 = \lambda_2$.

(i) Generate a value of U given $S_T = s_T$.

(ii) Generate the number of jump events N_T given U obtained in step (i).

(iii) Generate the event times $X_1, ..., X_{N_T}$ given U and N_T obtained in steps (i) and (ii). Then compute the whole path of volatility $\{v_t, 0 \le t \le T\}$.

(iv) Generate a path of prices $\{S_{t_i}, i = 1, ..., n-1\}$ given $\{v_t, 0 \le t \le T\}$ obtained in step (iii), and $S_{t_n} = S_T$.

6.1.1 Generate U given $S_T = s_T$

In this step, we sample the integral of variance U from the conditional distribution $f_{U|S_T}$ given $S_0 = s_0$, $S_T = s_T$ and $v_0 = \sigma_1$. By Bayes' theorem, this density is proportional to $(\propto)$ $f_{S_T|U} f_U$. Based on (6.2), $f_{S_T|U}$ has a log-normal distribution and can be represented as

$$f_{S_T|U}(S_T = s_T | U = u) = \frac{1}{z\sqrt{2\pi}\sqrt{u}} e^{-\frac{(\ln s_T - \ln s_0 - rT + \frac{1}{2}u)^2}{2u}}. \tag{6.10}$$

The characteristic function of U, ϕ_U, is given in (6.9), and hence we can use cosine series expansions to approximate f_U using ϕ_U by the way presented in Section 5.4.1. Then we can apply the inverse transform method by integrating numerically the density $f_{U|S_T}$. In our implementation, we use the method proposed by Chen et al. (2011), which we present in Appendix 6.A.

6.1.2 Generate N_T given U

In this step, we generate the number of events N_T during the interval $[0, T]$ given U obtained in Section 6.1.1. Hereafter, let $N \equiv N_T$. By Bayes' theorem, $f_{N|U} \propto f_{U|N} f_N$. We know that f_N follows a Poisson distribution with parameter λ_T, so the important part is to derive the distribution $f_{U|N}$. By the change of variables theorem, we have

$$f_{U|N}(U = u|N = n) \quad = \quad f_{T_1|N}\Big|\frac{\partial T_1}{\partial U}\Big| = \frac{1}{|\sigma_1^2 - \sigma_2^2|} f_{T_1|N}(T_1 = \frac{u - \sigma_2^2 T}{\sigma_1^2 - \sigma_2^2}|N = n).$$

Thus, the problem of finding $f_{U|N}$ is reduced to finding $f_{T_1|N}$.

Let $F_{T_1|N}$ denote the cumulative distribution function of T_1 given the total number of events N. In Proposition 6.1.1 bellow, we present a formula for $F_{T_1|N}$ when $N \leq 10$. Using a similar technique to the one we present in our proof, it is possible to extend our result to other values of N. However, in practice, it is often sufficient to consider only a finite range for N, as the probabilities of large values of N occurring are very small. For example, if $\lambda = 1.5$, then the probability of 10 events occurring over the interval $[0, 1]$ is less than 10^{-5}.

Proposition 6.1.1. For $N = 0$,

$$F_{T_1|N}(T_1 \leq l|N = 0) = \begin{cases} 1, & \text{if } l \geq T, \\ 0, & \text{otherwise.} \end{cases}$$

For any integer $n = 1, ..., 10$, we have

$$F_{T_1|N}(T_1 \leq l|N = n) = \sum_{j=0}^{\left\lceil \frac{n}{2} \right\rceil - 1} \binom{n}{j} (\frac{l}{T})^{n-j} (1 - \frac{l}{T})^j, \tag{6.11}$$

where $\binom{n}{j} = \frac{n!}{(n-j)!j!}$ and $\left\lceil \cdot \right\rceil$ represents the ceiling function.

Proof: See Appendix 6.B. $\square$

Remark 6.1.1. My committee member, Professor Don McLeish, suggested that the distribution in equation (6.11) is valid for all the positive integers. To prove this, it is equivalent to deriving the distribution of the median of the ordered statistics of i.i.d. random variables drawn from $U(0,T)$.

From the above result we can represent the densities $f_{T_1|N}(T_1 = l|N = n)$, where $n = 1, .., 10$, as follows:

$$f_{T_1|N}(T_1 = l|N = n) = \sum_{j=0}^{\lceil \frac{n}{2} \rceil - 1} \binom{n}{j} \Big(\frac{(n-j)l^{n-j-1}(T-l)^j}{T^n} - \frac{jl^{n-j}(T-l)^{j-1}}{T^n} \Big).$$

Therefore,

$$f_{N|U}(N = n|U = u) \propto \frac{\lambda_\Delta^n}{n!} f_{T_1|N}(T_1 = l|N = n), \text{ where } n = 1, ..., 10 \text{ and } l = \frac{u - \sigma_2^2 T}{\sigma_1^2 - \sigma_2^2}.$$

We normalize $f_{N|U}$ by the constant

$$D := \sum_{j=0}^{10} \frac{\lambda_\Delta^j}{j!} f_{T_1|N}(T_1 = l|N = j), \text{ where } l = \frac{u - \sigma_2^2 T}{\sigma_1^2 - \sigma_2^2}.$$

The above results give us all the components that are necessary to formulate an algorithm for generating the number of jumps N over the interval $[0,T]$ given U. We outline its steps in Algorithm 6.2. Note that sampling N from the distribution $f_{N|U}$ can be seen as sampling from a mixture model whose weights are defined by $\frac{1}{D} f_{T_1|N}$, where $N = 0, 1, ..., 10$.

Algorithm 6.2 (step (ii) of Algorithm 6.1.)

(i) Generate $u \sim U(0,1)$.

(ii) If $u \leq \frac{1}{D} f_{T_1|N}(T_1 = l|N = 0)$, return $N = 0$.

(iii) If $\frac{1}{D} \sum_{j=0}^{n} \frac{\lambda_\Delta^j}{j!} f_{T_1|N}(T_1 = l|N = j) \leq u < \frac{1}{D} \sum_{j=0}^{n+1} \frac{\lambda_\Delta^j}{j!} f_{T_1|N}(T_1 = l|N = j)$, $n = 1, ..., 9$, return $N = n$. Else, return $N = 10$.

6.1.3 Generate $X_1, ..., X_N$ given U and N

In this step, we need to generate $\underline{X} = \{X_1, ..., X_N\}$ from the conditional distribution $f_{\underline{X}|N,U}$, where N and U are obtained in Sections 6.1.1 and 6.1.2. The following proposition characterizes the distribution of $f_{\underline{X}|N,U}$.

Proposition 6.1.2. Given the number of jump events $N = n$ and the integral of variance $U = u$, the joint distribution of the jump times $\underline{X} = \{X_1, ..., X_n\}$ is uniform over the region $[0, T]$ where $0 < X_1 < ... < X_n < T$ and one of the following two conditions holds:

(i) If n is odd, then $\sum_{i=1}^{\frac{n+1}{2}}(X_{2i-1} - X_{2i-2}) = \frac{u-\sigma_2^2 T}{\sigma_1^2-\sigma_2^2}$.

(ii) If n is even, then $\sum_{i=1}^{\frac{n}{2}}(X_{2i} - X_{2i-1}) = \frac{u-\sigma_2^2 T}{\sigma_1^2-\sigma_2^2}$.

Proof: See Appendix 6.C. □

To describe a method of generating jump times based on the above results, we need to introduce some new notations.

(i) Denote the number of changes over $[0, T]$ from regime 1 to regime 2 by $N_{1,2}$, and the number of changes over $[0, T]$ from regime 2 to regime 1 by $N_{2,1}$, i.e., $N_{1,2}+N_{2,1} = N$. Given U, N and the initial state (which is 1 as assumed), we can determine $N_{1,2}$, $N_{2,1}$ and the total time spent in regime 1 and regime 2, T_1 and T_2.

(ii) Then we have $N_{1,2}+1$ time segments for regime 1 and $N_{2,1}$ time segments for regime 2, where by time segment we mean the time slot that the Markov chain stays in a single regime before it jumps to another regime.

(iii) Denote the time lengths of the segments for regime 1 by $Y_1, ..., Y_{N_{1,2}+1}$ and for regime 2 by $Z_1, ..., Z_{N_{2,1}}$. Then we have $Y_1 + ... + Y_{N_{1,2}+1} = T_1$ and $Z_1 + ... + Z_{N_{2,1}} = T_2$.

(iv) Note that given the time lengths of these segments, we can compute the jump times $\underline{X}$. For example, if $N = 4$ and hence $N_{1,2} = N_{2,1} = 2$, then we have $X_0 = 0, X_1 = Y_1, X_2 = Y_1 + Z_1, X_3 = Y_1 + Z_1 + Y_2, X_4 = Y_1 + Y_2 + Z_1 + Z_2$ and $X_5 = Y_1 + Z_1 + Y_2 + Z_2 + Y_3 = T$.

Thus the problem of generating jump times is reduced to the problem of deciding the time lengths of those segments. This means we need to find a way to split T_1 into $N_{1,2}+1$ parts and T_2 into $N_{2,1}$ parts. For convenience, we introduce the splitting times as follows, which can be represented by the time lengths of the segments defined above.

(v) Given T_1 and $N_{1,2}$, we need to determine the $N_{1,2}$ splitting times, $W_1 < ... < W_{N_{1,2}}$, over the interval $[0, T_1]$. Similarly, given T_2 and $N_{2,1}$, we need to determine the $N_{2,1}-1$ splitting times, $J_1 < ... < J_{N_{2,1}-1}$, over the interval $[0, T_2]$.

(vi) Once we have the splitting times, we can determine the time lengths of the segments. For example, if $N_{1,2} = 2$ and given W_1 and W_2, then we have $Y_1 = W_1, Y_2 = W_2 - W_1, Y_3 = T_1 - W_2$.

Let $\underline{W} = \{W_1, ..., W_{N_{1,2}}\}$ and $\underline{J} = \{J_1, ..., J_{N_{2,1}-1}\}$. The following lemma provides the distributions of $f_{\underline{W}|T_1,N_{1,2}}$ and $f_{\underline{J}|T_2,N_{2,1}}$.

Lemma 6.1.1. Given T_1 and $N_{1,2}$, the joint distribution of $\underline{W} = \{W_1, ..., W_{N_{1,2}}\}$ is uniform over the interval $[0, T_1]$ if $0 < W_1 < ... < W_{N_{1,2}} < T_1$. Similarly, given T_2 and $N_{2,1}$, the joint distribution of $\underline{J} = \{J_1, ..., J_{N_{2,1}-1}\}$ is uniform over the interval $[0, T_2]$ if $0 < J_1 < ... < J_{N_{2,1}-1} < T_2$.

Proof: See Appendix 6.D. □

After obtaining $\underline{W}$ and $\underline{J}$, we can compute the time lengths of the segments as described above and hence the event times $\underline{X}$. Algorithm 6.3 summarizes the steps to generate $X_1, ..., X_N$ given U and N.

Algorithm 6.3 (step (iii) of Algorithm 6.1.)

Require: Given U and N.

Output: Jump times $X_1, ..., X_N$.

(i) Compute T_1, T_2, N_1 and N_2 given U and N.

(ii) Uniformly generate $W_1 < ... < W_{N_{1,2}}$ from the interval $[0, T_1]$ and $J_1 < ... < J_{N_{2,1}-1}$ from the interval $[0, T_2]$ respectively.

(iii) Obtain $X_1, ..., X_N$ by using $W_1 < ... < W_{N_{1,2}}$ and $J_1 < ... < J_{N_{2,1}-1}$.

6.1.4 Generate $\{S_{t_i}, i = 1, ..., n\}$ given U, N and $X_1, ..., X_N$

In this step, we generate a path of $\{S_{t_i}, i = 1, ..., n\}$ given $S_0 = s_0, S_T = s_T$ and all the information obtained in Sections 6.1.1 – 6.1.3, i.e., we would like to sample S_t from the conditional distribution $f(S_t|S_0, S_T)$, where $t \in \{t_i, i =, ..., n\}$.

By Bayes' theorem, we have

$$f(S_t|S_0, S_T) = \frac{f(S_t, S_0, S_T)}{f(S_0, S_T)} = \frac{f(S_T|S_t)f(S_t|S_0)}{f(S_T|S_0)} \propto f(S_T|S_t)f(S_t|S_0) \tag{6.12}$$

Based on the results in (6.2) and (6.3), we know that $f(S_T|S_t)$, $f(S_t|S_0)$ and $f(S_T|S_0)$ in equation (6.12) all follow log-normal distributions, and hence (6.12) is known explicitly. Through a simple algebra, we can show that $f(S_t|S_0, S_T)$ is proportional to a log-normal distribution with known mean and variance (details are presented in Appendix 6.F). Then for any $t \in \{t_i, i = 1, ..., n\}$, we can sample $\{S_t\}$ from this distribution directly.

6.1.5 Algorithm when $\lambda_1 \neq \lambda_2$

In this section, we show that we can still use Algorithm 6.1 when $\lambda_1 \neq \lambda_2$, but for this case the algorithm must be augmented with an acceptance-rejection step. It is known that the likelihood function of a Markov chain with the infinitesimal generator A defined in (6.4) is of the form (e.g., Bladt and Srensen (2005))

$$L_1 := \lambda_1^{N_{1,2}} e^{-\lambda_1 T_1} \lambda_2^{N_{2,1}} e^{-\lambda_2 T_2}. \tag{6.13}$$

When $\lambda_1 = \lambda_2 = \lambda$, we have

$$L_0 := \lambda^N e^{-\lambda T}. \tag{6.14}$$

To apply the rejection sampling method, we need to find a constant $C_L > 1$ such that $L_1 \leq C_L L_0$, which means we need to find the upper bound on $\frac{L_1}{L_0}$. Note that when $N = 0$ and $L_1 = L_0$, then $L_1 \leq C_L L_0$ holds for any $C_L > 1$. Thus, we only consider the case when $N \neq 0$.

Define

$$L_{1,0} := \frac{L_1}{L_0} = (\frac{\lambda_2}{\lambda_1})^{N_{2,1}} e^{-(\lambda_2 - \lambda_1)T_2}. \tag{6.15}$$

To find the upper bound of (6.15), we consider two cases:

(i) $\lambda_1 > \lambda_2$.

In this case, we have $0 < \frac{\lambda_2}{\lambda_1} < 1$ and $\lambda_2 - \lambda_1 < 0$, and hence

$$L_{1,0} < 1 \cdot e^{(\lambda_1 - \lambda_2)T_2} \leq e^{(\lambda_1 - \lambda_2)T}.$$

Thus we take $C_L = e^{(\lambda_1 - \lambda_2)T}$.

(ii) $\lambda_1 < \lambda_2$.

In this case, we have $\frac{\lambda_2}{\lambda_1} > 1$ and $\lambda_2 - \lambda_1 > 0$, and hence

$$L_{1,0} < \frac{\lambda_2}{\lambda_1}^{N_{2,1}} \cdot 1 \leq (\frac{\lambda_2}{\lambda_1})^N.$$

Thus we take $C_L = (\frac{\lambda_2}{\lambda_1})^N$.

In the following algorithm, Algorithm 6.4, we summarize the steps to generate a path of $\{S_{t_i}, i = 1, ..., n\}$ given $S_0 = s_0$ and $S_T = s_T$ when $\lambda_1 \neq \lambda_2$.

Algorithm 6.4

Require: Given $S_0 = s_0$, $S_T = s_T$, $v_0 = \sigma_1$, $\Delta = T$, the infinitesimal generator of the Markov chain, $\lambda_1 = c_1$ and $\lambda_2 = c_2$, where $c_1 \neq c_2$. Compute C_L using c_1 and c_2.

Output: A path of $\{S_{t_i}, i = 1, ..., n\}$ under the assumption that $\lambda_1 \neq \lambda_2$.

(i) Repeat

(i-1) Generate $v \sim U(0,1)$.

(i-2) Set $\lambda_2 = \lambda_1 = c_1$. Then use Algorithm 6.1 to obtain U, N and a path of $\{v_t, 0 \leq t \leq T\}$, and then compute L_0.

(i-3) Use $\lambda_1 = c_1, \lambda_2 = c_2$ and U, N obtained in step (i-2) to compute L_1.

until $v < \frac{L_0}{C_L L_1}$.

(ii) Generate a path of $\{S_{t_i}, i = 1, ..., n\}$ given $\{v_t, 0 \leq t \leq T\}$ obtained in step (i-2).

6.2 Pricing Asian Options

In this section, we apply Algorithm 6.4 (or Algorithm 6.1, depending on model parameters) to the problem of pricing Asian options for the regime switching Black-Scholes model defined in (6.1).

Denote the payoff of a path-dependent option at time $t \in [0, T]$ by

$$C(S, t) \equiv C(S_u, t)_{u \in [0,t]}.$$

Here we only consider an Asian call option on arithmetic average with fixed strike. Then the payoff of the Asian option at maturity T is given by

$$C_{AF}(S, T) := (A_n - K)^+, \tag{6.16}$$

where K is the strike price, $A_n = \frac{1}{n}\sum_{i=1}^{n} S_{t_i}$ and $\{t_1, ..., t_n = T\}$ is a set of equally spaced monitoring dates. For model (6.1), the arbitrage-free price of the Asian option at time t can be represented in the following form:

$$V(S, t) := e^{-r(T-t)} E^Q[C_{AF}(S, T)|\mathcal{F}_t], \tag{6.17}$$

where Q is a risk-neutral measure, r is a continuously compounded interest rate and is assumed to be a constant, $\{\mathcal{F}_t\}$ is a natural filtration generated by the process $\{S_t\}$. We can obtain (6.17) by the crude Monte Carlo simulation method, where we need to sample the whole path of the price process $\{S_{t_i}, i = 1, ..., n\}$. For convenience, we assume that the number of steps n is a power of two. In addition, let M be the number of paths need to be generated and N_{MC} be the number of repetitions for the Monte Carlo simulation.

For comparison, we propose two ways to sample paths and then price the Asian option, and we refer to them as the forward method and the sampling method for the conditioned process. The main steps of these two methods are outlined below.

(i) Forward Method

For $j = 1, ..., M$ repeat (S5–1)–(S5–6):

(S5–1) Generate the number of event $N \sim POI(\lambda_T)$.

(S5–2) Generate the event times $0 < X_1 < ... < X_N < T$ uniformly on the interval $[0, T]$. Then we have the whole path of $\{v_s, 0 \leq s \leq T\}$.

(S5–3) Compute U_{t_{i-1},t_i} by (6.5), $i = 1, ..., N$.

(S5–4) Generate $L_{\Delta,i} \sim N(r(t_i - t_{i-1}) - \frac{1}{2}U_{t_{i-1},t_i}, U_{t_{i-1},t_i}), i = 1, ..., n$.

(S5–5) Obtain a path of $\{S_{t_i}\}$ by computing the equation $S_{t_i} = S_{t_{i-1}} e^{L_{\Delta,i}}, i = 1, ..., n$.

(S5–6) Compute the discounted payoff at time 0 for the j-th path as $V_j = e^{-rT}(\frac{1}{n}\sum_{i=1}^{n} S_{t_i} - K)^+$.

(S5–7) Average the M discounted payoffs, then we have the price of the option at time 0 as $\frac{1}{M}\sum_{j=1}^{M} V_j$.

(ii) Sampling Method for the Conditioned Process

For $j = 1, ..., M$ repeat (S6–1)–(S6–6):

(S6–1) Generate the number of event $N \sim POI(\lambda_T)$.

(S6–2) Generate event times $0 < X_1 < ... < X_N < T$ uniformly on the interval $[0, T]$.

(S6–3) Compute the time spent on regime 1 and regime 2 as T_1 and T_2 given the event times generated in (S6–2). Then we obtain $U = \sigma_1^2 T_1 + \sigma_2^2 T_2$.

(S6–4) Sample $L_T \sim N(rT - \frac{1}{2}U, U)$. Obtain S_T by computing $S_T = S_0 e^{L_T}$, where S_0 is given.

(S6–5) Generate a path of $\{S_{t_i}\}$ by Algorithm 6.4 (or Algorithm 6.1) given S_T that is obtained in (S5–4).

(S6–6) Compute the discounted payoff at time 0 for the j-th path as $V_j = e^{-rT}(\frac{1}{n}\sum_{i=1}^{n} S_{t_i} - K)^+$.

(S6–7) Average the M discounted payoffs, then we have the price of the option at time 0 as $\frac{1}{M}\sum_{j=1}^{M} V_j$.

Next, we provide an example to illustrate the above two methods to price the Asian option described in this section under the two-regime switching Black-Scholes model defined in (6.1). In our implementation, we have used the same model parameters as those used in Fuh et al. (2012), they are $\sigma_1 = 0.2, \sigma_2 = 0.3, \lambda_1 = \lambda_2 = 10, r = 0.1$. The other parameters that we use are $T = 1, n = 256, M = 500, K = 1$ and $S_0 = 1$. Because in this example $\lambda_1 = \lambda_2$, we use Algorithm 6.1 to sample conditioned processes from the model.

Table 6.1 shows the estimated prices of the Asian option at time 0 and the computational time (CPU time in seconds) obtained by the forward method and the sampling method for the conditioned process (SMCP) based on $N_{MC} = 50$ repetitions. The numbers in the parentheses are the standard deviations of the estimates. We can see that the estimates and their standard deviations are similar for these two methods, but the forward method is more efficient. Indeed, the most time consuming part of the sampling method for the conditioned process is step (S6–5). We also repeated our calculations for other selections of the model parameters, and the results, in terms of speed and accuracy, are very similar to those reported in Table 6.1. To improve the computational efficiency of the SMCP, we can consider an efficient method of filtering, which is further discussed in Section 6.4, and hence we do not need to generate different paths of regime for each repetition of the sampling algorithm.

Table 6.1: Comparison of Two Sampling Methods for an Asian Option

Method	Price	CPU (in seconds)
Forward	0.0783(0.005)	104.3
SMCP	0.0737(0.005)	270.0

6.3 Static Hedging of Asian Options

In this section, we study the problem of static hedging of path-dependent options by using European options under a two-regime switching Black-Scholes model. This static hedging problem has been solved under the Black-Scholes model by Kolkiewicz (2016). In the paper, the author proposes the optimal hedging strategy that minimizes the shortfall risk and compares it with the one that minimizes the mean squared hedging error.

There are several differences between our optimal hedging problem and the one considered by Kolkiewicz (2016). First, the optimal hedging options under a regime switching model depend not only on the path of prices but also on the path of regime. Second, generating paths of prices given terminal values under our model is no longer equivalent to generating Brownian bridges as described in the paper. The theoretical results in that paper can be relatively easily extended to our models, which we present in Section 6.3.1. However, the main challenge is in the implementation, where we need to apply Algorithm 6.1 to sample paths of the underlying asset. In Section 6.3.2, we consider the same Asian option described in Section 6.2 and we use the proposed method in Section 6.3.1 to find the optimal hedging strategies of the Asian option under our models.

First, we formulate our models and introduce some definitions and notations as follows. Assume that under the real-world probability measure P, the dynamics of the process $\{S_t\}$ is given by

$$dS_t = \theta_t S_t dt + v_t S_t dW_t, t \in [0, T], \tag{6.18}$$

where the parameters $\{\theta_t\}$ and $\{v_t\}$ are independent of the standard Brownian motion processes $\{W_t\}$ and follow a continuous-time Markov chain. Let $\{Y_t\}$ be the process of the regime. Assume there are two states, i.e., if $Y_t = i$, then $\theta_t = \mu_i$ and $v_t = \sigma_i$, where $i = 1, 2$. conditional on the paths of $\{\theta_t\}$ and $\{v_t\}$, we have

$$\ln \frac{S_t}{S_0} \sim N\Big(\int_0^t \theta_s ds - \frac{1}{2} \int_0^t v_s^2 ds, \int_0^t v_s^2 ds \Big), \quad t \in [0, T]. \tag{6.19}$$

To describe a practical discrete-time hedging problem, let us consider a single time interval starting from $t = 0$. Suppose that we want to hedge a path-dependent option over a time interval of length T_h, where $T_h \in [0, T]$. The objective is to create a static portfolio including European options (calls and/or puts with different strike prices) on the underlying asset and a bank account so that its value at T_h is as close as possible to the value of the path-dependent option $C(S, T_h)$. The initial cost of the portfolio is assumed to not exceed a given budget, say V_I. For a certain function h, $h(S_{T_h})$ can represent such a hedging strategy. Define the hedging error as $C(S, T_h) - h(S_{T_h})$ and the shortfall risk as $\big(C(S, T_h) - h(S_{T_h})\big)^+$.

Since our objective is to find an optimal hedging option h so that the hedging error meets our risk management objective, it is more convenient for us to use an alternative description. Let us assume that $S_{T_h} = s$ is given and $\{S_{t|s}\}$ represents the process $\{S_t\}$ conditioned on $S_{T_h} = s$. Then the payoff function of a path-dependent option can be written in a form where the terminal value s and the conditional process $\{S_{t|s}\}$ are separated.

conditional on $S_{T_h} = s$, the risk of hedging $C(S, T_h)$ by $h(S_{T_h})$ depends only on the random variable

$$L(s, T_h) := C(S_{t|s}, T_h)_{t \in [0, T_h]},$$

which we will call the conditional residual risk. Therefore, we can use the set of random variables

$$\mathcal{L}(T_h) := \{L(s, T_h) : s \in \mathbb{R}^+\}$$

and the distribution of the terminal price S_{T_h} to completely describe the hedging error. For simplicity, we shall denote $L(s, T_h)$ by $L(s)$.

6.3.1 Optimal Hedging Strategies

Once the residual risk is known, we can hedge the option by selecting a European option with maturity T_h such that a particular risk management objective is satisfied. Such strategy can be represented by finding the value of a specified function, say $h(s)$[2], $s \in \mathbb{R}^+$, such that the hedging error or the shortfall risk has desirable properties.

In this section, we discuss the same two optimal hedging strategies as presented in Kolkiewicz (2016). They are obtained by respectively minimizing the mean squared hedging error and the shortfall risk. Now we describe the two optimal hedging options under our models. Let

$$\mathcal{S}_L := \{s \in \mathbb{R}^+ : \text{interior of } supp(L(s)) \text{ is non-empty}\},$$

where $supp(\cdot)$ represents the support of a random variable and

$$\mathcal{H}^0 := \{\text{functions } h \text{ on } \mathcal{S}_L : h(s) \in \text{ the closure of } supp(L(s)) \text{ for } s \in \mathcal{S}_L\}.$$

Assume that the set $supp(L(s))$ is connected [3] for each $s \in \mathcal{S}_L$. Given the initial capital V_I, define the set of possible hedging options h as

$$\mathcal{H} := \{h \in \mathcal{H}^0 : E^Q[h(S_{T_h})] \le V_0\},$$

where Q is a risk-neutral measure and $V_0 = e^{rT_h} V_I$.

[2]The subscript h in T_h is only a notation, which is different from the function $h(\cdot)$.

[3]Note that $L(s), s \in \mathcal{S}_{\mathcal{L}}$, is a real-valued function, and it can be seen as the payoff function of the underlying asset with $S_{T_h} = s$. The set $supp(L(s))$ is connected for each $s \in \mathcal{S}_{\mathcal{L}}$, and hence they form a metric topology.

Define $\mathcal{L}^2(S_{T_h})$ as the set of measurable and square integrable functions of S_{T_h}. Then the optimal hedging option that minimizes the mean-square of hedging error under the measure P is defined as

$$h_e := \arg \inf_{h \in \mathcal{H}_2} E^P\big[\big(C(S,T_h) - h(S_{T_h})\big)^2\big], \tag{6.20}$$

where $\mathcal{H}_2 := \mathcal{H} \cap \mathcal{L}^2(S_{T_h})$. As shown in Kolkiewicz and Liu (2012), we have an analytical representation of h_e:

$$h_e(s) := E^P\big[C(S,T_h)|S_{T_h} = s\big]. \tag{6.21}$$

It is easy to verify that $E^P\big[h_e(s)\big] = V_0$.

Now we consider the optimal hedging strategy that minimizes the shortfall risk, denoted by h_{opt}. Define

$$\mathcal{H}_{h_L,h_U} := \{h \in \mathcal{H} : h_L(s) \leq h(s) \leq h_U(s), s \in \mathcal{S}_L\},$$

where the given functions h_L and h_U satisfy the following assumptions:

(A4–1) h_L and h_U belong to H_0 and are continuous.

(A4–2) $h_L(s) < h_U(s)$, where $s \in supp(S_{T_h}) \cap \mathcal{S}_L$.

(A4–3) $E^Q[h_L(S_{T_h})] \leq V_0$ and $V_0 \leq E^Q[h_U(S_{T_h})] < \infty$.

(A4–4) $E^P[(h_U(S_{T_h}) - h_L(S_{T_h}))^p] < 1$, where $p \geq 1$ is a pre-determined value[4].

Define the optimal hedging option h_{opt} as

$$h_{opt} := \arg \inf_{h \in \mathcal{H}_{h_L,h_U}} E^P\left[\Big(\big(C(S,T_h) - h(S_{T_h})\big)^+\Big)^p\right]. \tag{6.22}$$

The solution to the above optimization problem can be represented in terms of the following function

$$g(s,z;p) := \frac{g_0(s, z(h_U(s) - h_L(s)) - h_U(s))}{(h_U(s) - h_L(s))^p}, \ (s,z) \in \mathcal{S}_L \times [0,1], \tag{6.23}$$

where

$$g_0(s,z;p) := E^P\big[\big((L(s)+z)^+\big)^p\big], \ (s,z) \in \mathcal{S}_L \times [-h_U(s), -h_L(s)]. \tag{6.24}$$

Note that $z \to g(s,z;p), z \in [0,1]$, is convex and non-decreasing. We also assume that

[4]We use $p = 2$ in our implementation, because we would like to make h_{opt} and h_e comparable.

(A4–5) For each $s \in \mathcal{S}_L$, the function $z \to g(s,z;p)$ is continuously differentiable at any $z \in [0,1]$. We consider one-sided derivatives at the end points $z=0$ and $z=1$, and its derivative $g_z(s,z;p) := \frac{\partial g(s,z;p)}{\partial z}$ is strictly increasing.

(A4–6) $E^P\left[\left(h_U(S_{T_h}) - h_L(S_{T_h})\right) g_z(S_{T_h}, -h_L(S_{T_h});p)\right] < \infty$.

Denote the inverse of the function $z \to g_z(s,z;p)$ by $l_g(s,y)$, $y \in [g_z(s,0;p), g_z(s,1;p)]$, and assume that it is well defined. Define the extended inverse function as

$$l_e(s,y) = \begin{cases} l_g(s,y), & \text{if } y \in [g_z(s,0;p), g_z(s,1;p)], \\ 0, & \text{if } y < g_z(s,0;p), \\ 1, & \text{if } y > g_z(s,1;p). \end{cases}$$

Let P^* and Q^* be the distributions of S_{T_h} under the measures P and Q respectively. Then conditional on the volatility path, these two distributions of S_{T_h} can be represented as log-normal distributions as respectively described in (6.2) and (6.19).

Before presenting the solution to the optimization problem in (6.22), we need to identify the form of $\frac{dQ^*}{dP^*}$ under a two-regime switching model, which is needed in the representation of the solution. Based on the model assumptions, $\int_0^{T_h} \theta_s ds$ and $\int_0^{T_h} v_s^2 ds$ can be represented in terms of the occupation time in regime 1. Let T_1 represent the time spent in regime 1 over the time interval $[0, T_h]$, then $\int_0^{T_h} \theta_s ds = \mu_1 T_1 + \mu_2(T_h - T_1)$ and $\int_0^{T_h} v_s ds = \sigma_1^2 T_1 + \sigma_2^2(T_h - T_1)$. Define $\bar{\mu}(T_1) := \int_0^{T_h} \theta_s ds$ and $\bar{\sigma}(T_1) := \int_0^{T_h} v_s^2 ds$. Denote the density function of T_1 by f_{T_1}. Then we have

$$\frac{dQ^*}{dP^*}(s) \equiv \frac{dQ^*(s)}{dP^*(s)} := \frac{\int_0^{T_h} \frac{1}{s\sqrt{2\pi\bar{\sigma}(t)}} \exp(-\frac{(\ln s - \ln S_0 - rT_h + \frac{1}{2}\bar{\sigma}(t))^2}{2\bar{\sigma}(t)}) f_{T_1}(T_1 = t) dt}{\int_0^{T_h} \frac{1}{s\sqrt{2\pi\bar{\sigma}(t)}} \exp(-\frac{(\ln s - \ln S_0 - \bar{\mu}(t) + \frac{1}{2}\bar{\sigma}(t))^2}{2\bar{\sigma}(t)}) f_{T_1}(T_1 = t) dt}, \quad s \in \mathcal{S}_L. \tag{6.25}$$

Although the density function of T_1, f_{T_1}, is unknown explicitly, the characteristic function of T_1, ϕ_{T_1}, is given in (6.7) with $T = T_h$. Again we can use cosine series expansions to approximate f_{T_1} by using ϕ_{T_1}.

The following proposition provides the solution to the optimization problem stated in (6.22).

Proposition 6.3.1. Suppose that the assumptions (A4-1)–(A4-6) hold. Then the solution of the optimization problem (6.22) can be written in the form of

$$h_{opt}(s) = h_L(s) + \bar{\gamma}(s)\big[h_U(s) - h_L(s)\big],\ s \in \mathcal{S}_L, \tag{6.26}$$

where

$$\bar{\gamma}(s) = 1 - l_e\Big(s, c \cdot \big[h_U(s) - h_L(s)\big]^{1-p} \frac{dQ^*}{dP^*}(s)\Big). \tag{6.27}$$

The constant c is selected by solving the following equation

$$E^{Q^*}\big[h_{opt}(S_{T_h})\big] = V_0. \tag{6.28}$$

This constant c exists and is unique.

Proof: By Kolkiewcz's proof, it suffices to show that the likelihood ratio $\frac{dQ^*}{dP^*}$ is continuous for our specification of the measures P and Q. See the proof in Appendix 6.E. □

6.3.2 Implementation

In this section, we provide a numerical example to illustrate the hedging strategies presented in Section 6.3.1. We use a Monte Carlo simulation method to obtain the optimal hedging options h_e and h_{opt} for the Asian option that is used in Section 6.2. To generate paths from the conditioned processes under a two-regime switching model[5], we need to apply Algorithm 6.1 proposed in Section 6.1. In Kolkiewicz (2016), the conditional residual risk of an option can be represented using a Brownian bridge, which makes the implementation easier. However, this simpler form of the conditional residual risk is not available in our models. Given $S_{T_h} = s$, the conditional residual risk of the Asian option can be written as

$$L_{AF}(s, T_h) := (A_n^s - K)^+,\ s \in \mathbb{R}^+,$$

with

$$A_n^s = \frac{1}{n}\Big(\sum_{i=1}^{n-1} S_{t_i|s} + S_{t_n}\Big) = \frac{1}{n}\Big(\sum_{i=1}^{n-1} S_{t_i|s} + s\Big).$$

[5]We only consider the two-regime switching Black-Scholes model in this section, since we have only proposed the sampling algorithm for this model. For other regime switching models, we would like to propose algorithms to sample the conditioned processes in future research.

Then our problem now is to find a way to generate the process $\{S_{t|s}\}$ given $S_{T_h} = s$, which is the same problem that has been solved by Algorithm 6.1 in Section 6.1. More details are provided later in this section.

To sample a path of $\{S_t\}$ given $S_T = s_T$ under model (6.18), we can apply Algorithm 6.1 with some modifications. First, under model (6.18), the distribution of S_T given U in step (i) of Algorithm 6.1 is no longer in the form of (6.10) and it becomes

$$f_{S_T|U}(S_T = x|U = u) = \frac{1}{x\sqrt{2\pi u}} e^{-\frac{(\ln x - \ln S_0 - \bar{\theta}(u) + \frac{1}{2}u)^2}{2u}}, \tag{6.29}$$

where

$$\bar{\theta}(u) = \mu_1 \frac{u - \sigma_2^2 T}{\sigma_1^2 - \sigma_2^2} - \mu_2 \frac{u - \sigma_1^2 T}{\sigma_1^2 - \sigma_2^2}.$$

Second, in step (iii) of Algorithm 6.1, we need to compute not only the path of $\{v_t\}$ but also the path of $\{\theta_t\}$. Finally, we need to replace rt and $r(T-t)$ with $\int_0^t \theta_s ds$ and $\int_t^T \theta_s ds$ respectively in the log-normal distribution derived in Appendix 6.F to sample $\{S_t\}$.

In our implementation, for pre-determined levels α_L and α_H such that $0 < \alpha_L < \alpha_H < 1$, we define h_L and h_U as the α_L-quantile and α_H-quantile of the conditional residual risk of the option.

The main steps for obtaining the optimal hedging options proposed in (6.21) and (6.26) of the Asian option under a two-regime switching model are outlined as below.

For $T_h = T$, select an equally spaced set of positive points $\mathcal{S} := \{s_1, ..., s_{N_s}\}$ such that $P(S_T \in [s_1, s_{N_s}]) = \alpha_s$, where α_s is at least equal to a predetermined level, say 0.9999.

(S7-1) Select an equally spaced set of points from the interval $[0, T]$ as $u_{T_1} := \{u_1, ..., u_{n_t}\}$. Then use the trapezoidal rule with respect to u_{T_1} to calculate the numerator and denominator of (6.25) and store the resulting values as $dQ^*(s_j)$ and $dP^*(s_j)$ respectively for each $s_j,\ j = 1, ..., N_s$.

For each $s_j,\ j = 1, ..., N_s$, repeat (S7-2)–(S7-5):

(S7-2) Use Algorithm 6.1[6] with the modifications described above to obtain N_B independently sampled paths of $\{S_{t_i|s_j},\ i = 1, ..., n\}$. Then for each path $l,\ l = 1, ..., N_B$, compute the value of $A_n^{s_j}$ and denote it by a_{jl}. Then approximate the value of the

[6] To improve the efficiency, we refer to Section 6.4 for more details.

payoff function of the mean squared optimal hedging option h_e by the sample means of the payoff function as

$$h_e(s_j) \approx \frac{1}{N_B}\sum_{l=1}^{N_B}(a_{jl}-K)^+,\ j=1,...,N_s.$$

(S7-3) For pre-determined levels α_L and α_H such that $0<\alpha_L<\alpha_H<1$, obtain the lower and upper bounding functions $h_L(s_j)$ and $h_U(s_j)$ as the α_L-quantile and α_H-quantile of $(A_n^{s_j}-K)^+$ by using the samples $a_{jl},\ l=1,...,N_B$.

(S7-4) Select an equally spaced set of points $\mathcal{Z}:=\{z_1,...,z_{N_z}\}$ from the interval $[0,1]$. For each point, use a sample mean based on N_B independent paths generated in (S7-2) to approximate the value of function $g(z_k;s_j,p)$ defined in (6.23).

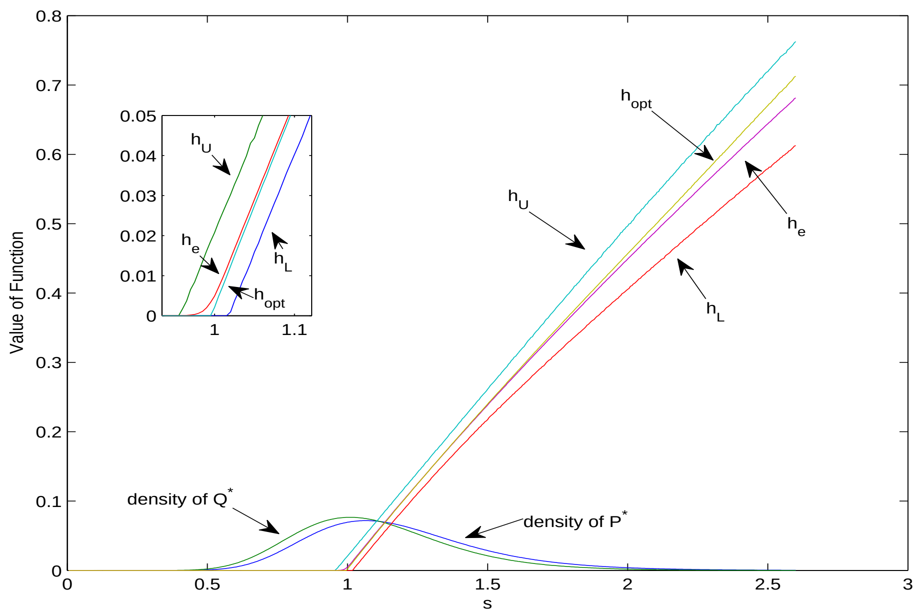
(S7-5) Approximate the derivatives $\frac{\partial g(z;s_j,p)}{\partial z}$ by using the finite difference method, where $z_k\in\mathcal{Z}$. Then find the corresponding inverse function l_g and the extended inverse l_e by the inverse transform method proposed by Chen et al. (2011) (presented in Appendix 6.A). Then approximate $\bar{\gamma}(s_j)$ defined in (6.27) by

$$\bar{\gamma}(s_j)\approx 1-l_e\Big(s_j,c\big[h_U(s_j)-h_L(s_j)\big]^{1-p}\frac{dQ^*(s_j)}{dP^*(s_j)}\Big).$$

(S7-6) By using $h_L(s),h_U(s)$ and l_e obtained in (S7-3) and (S7-5), where $s\in\mathcal{S}_L$, we can obtain the optimal hedging option $h_{opt}(s_j)$ defined in (6.26) up to the constant c. To determine c that is consistent with the budget constraint defined in equation (6.28), we use the Matlab build-in function 'fminsearch' with $h_{opt}(s_j)$ and $dQ^*(s_j),\ j=1,...,N_s$, that are obtained in previous steps.

We use the following parameters: $\mu_1=0.3,\mu_2=0.5,n=256,N_s=599,N_z=501,N_B=20000,p=1,s_1=0.005,s_{N_s}=3,\alpha_s=0.9999,\alpha_L=0.01,\alpha_H=0.99$ and other model parameters are the same as those used in Section 6.2. To make the optimal hedging options h_e and h_{opt} comparable, we choose V_I to be the price of the option at time 0. The resulting c is equal to 0.5653, and for this c the difference between the prices of the options h_e and h_{opt} is less then 10^{-6}. Figure 6.2 shows the resulting hedging options h_e and h_{opt}, where the y-axis is the value of these options and the x-axis is the terminal value of the price process $s\in\mathcal{S}_L$. The window in the figure is a zoomed part of the plot. We can see that these two hedging options are close. The expected shortfall for the mean squared optimal option h_e is 0.0021, while for the optimal option h_{opt} is 0.0020, giving a reduction by around 3.2%.

Figure 6.2: The mean squared and the optimal static hedging options for the Asian call option



6.4 Filtering Regimes

In this section, we briefly describe another application of Algorithm 6.1, where the model is defined in (6.18). The problem that we consider is the following: given the prices of the underlying asset at times 0 and T, we want to determine a likely path of volatility or a likely path of regime over the interval $[0, T]$.

Given $S_T = s_T$, we can sample U, an integral of variance over the interval $[0, T]$, by step (i) of Algorithm 6.1. However, there are many possibilities of the regime path over $[0, T]$ given U. We can sample these different paths of regime by steps (ii)–(iii) of Algorithm 6.1, but our question is which path is the right one. Intuitively, we can solve this problem by finding the path that has the highest probability to occur. In detail, we can obtain the likelihood functions of the sampled paths of the hidden Markov process and then find the path with the largest likelihood function. Let $\{y_t\}$ be the realization of the hidden Markov process $\{Y_t\}$. Then for a set of pre-determined times $0 = t_0 < t_1 ... < t_{n-1} < t_n = T$, the likelihood function of a sampled path of $\{Y_t\}$ is in the form of

$$L_Y := \prod_{i=1}^{n} P(Y_{t_i} = y_{t_i} | Y_{t_{i-1}} = y_{t_{i-1}}), \tag{6.30}$$

which is a product of transition probabilities of the Markov chain. For example, if we have N_Y, say 20,000, sampled paths of $\{Y_t\}$ whose likelihood are denoted by $L_Y^j,\ j = 1, ..., N_Y$. Let $L_{\max} := \max\{L_Y^j,\ j = 1, ..., N_Y\}$, then we choose the path of regime whose likelihood is equal to $L_{\max}$ as the selected right path of regime or the filtered path of regime.

Remark 6.4.1. As suggested by my committee member, Professor Tony Wirjanto, given the infinitesimal generator of the Markov chain that contains two regimes, we can derive the transition probabilities in equation (6.30) explicitly (see, for example, Kac (1974) and López and Ratanov (2014)). For a Markov chain that contains more than two regimes, we can consider several ways to approximate the transition densities, for example, we can derive the partial differential equations for the densities and solve them numerically.

6.5 Conclusions of Chapter 6

In this chapter, we propose a sampling algorithm to generate conditioned processes from a two-regime switching Black-Scholes model. Then we apply the algorithm to the problems of pricing and static hedging of Asian options. We extend the static hedging method for path-dependent options under the Black-Schole model proposed by Kolkiewicz (2016) to our models. All the numerical results show that our proposed sampling algorithm works well. In addition, to improve the efficiency of the proposed sampling algorithm, we would like to propose an efficient method of filtering. More work on this problem is left for future research.

Chapter 7

Future Research

In this chapter, we provide topics for future research.

7.1 Efficiency of the Selected Points for the DECF Method

In Chapter 2, we estimate model parameters by the DECF method, and the points at which a characteristic function is evaluated can be obtained by an application of the quantization method proposed by Pagés et al. (2004). In our approach, we search for a random vector taking values in $\Gamma = \{r_1, ..., r_q\}$ that minimizes $\int \min_{1 \leq i \leq q} |\mathbf{z} - r_i|^p dG_N(\mathbf{z})$, where $\mathbf{z}$ is uniformly distributed. Alternatively, we consider G_N to be the empirical characteristic function of observations. Although these approaches seem to produce estimators with reasonably good properties, it is not clear how they can be related more formally to the efficiency of the estimators. We would like to study different properties of the resulting estimators. In particular, we want to establish the rate at which $N \cdot var(\hat{\underline{\xi}})$ converges to its limit as described in Section 2.3.2.

We will also consider some extensions of the testing methods presented in Chapter 3:

(i) The formal testing procedure that we discuss in Section 3.2.2 also requires a proper selection of the points at which a characteristic function is evaluated, and hence it would be interesting to develop systematic ways of doing this. However, we need to use the power of the test instead of the efficiency as a criterion.

(ii) For the proposed goodness-of-fit test in Section 3.2.2, we only consider the case when $m = 1$ in the simulation study in Section 3.3. Since we use $m = 2$ for the estimation method, we would also like to extend the goodness-of-fit test to the case when $m = 2$.

(iii) Chen et al. (2013) propose an empirical likelihood approach that is based on conditional characteristic functions for estimation and testing of Markov models in Lévy processes. Naturally, we would like to extend their methods to regime switching models in Lévy processes, where we can use joint characteristic functions instead of conditional ones.

7.2 Pricing under Regime Switching Models

At the end of Chapter 5, we find that two exercise regions may occur when pricing ratchet EIAs with surrender risk. To price ratchet EIAs, we can use the COS method where Fang and Oosterlee (2008, 2009) propose a recursive way of calculating coefficients in Fourier expansions, which significantly reduces computational time. However, this technique requires finding the boundary that separates the exercise region from non-exercise region at each time step, usually by finding the root of a non-linear equation numerically. The same approach can also be applied to the PV method where we combine trigonometric functions with polynomials. Therefore, finding the possible two boundaries when pricing ratchet EIAs is essential if we want to reduce computational time and improve the convergence rate of prices.

Pricing perpetual ratchet EIAs is also an interesting problem, since it can be seen as pricing contracts with extreme long term maturities.

In addition, to price an early exercise feature under regime switching models described in terms of their characteristic functions, the existing methods assume that the current regime is known, which is not true in reality. Typically, we estimate the regime using either a calibration procedure or some filtering method. Rambharat (2012) shows how to price American options in a limited information framework, where the volatility of the underlying asset is a latent stochastic process. However, the author uses particle filter methods, an application of which requires knowledge of density functions. We would like to develop an efficient method of filtering for models where only characteristic functions are known.

7.3 Sampling Conditioned Processes

We considered some extensions of the proposed conditioned process sampling algorithm in Chapter 6:

(i) It is natural to extend Algorithm 6.1 to other regime switching models as formulated in Section 1.1.2. Compared with the regime switching Black-Scholes model, these models have more parameters that can change according to the hidden Markov process. Therefore, it is more reasonable to introduce more regimes rather than two as considered in Chapter 6.

(ii) To improve the computational efficiency of the conditioned process sampling algorithm used in the pricing and static hedging problems, we can find several paths of regime and then reuse them in the sampling algorithm rather than generating different paths of regime for each repetition of the algorithm. To solve this problem, we can use the idea of filtering presented in Section 6.4. For example, given $U = u$, we use the method proposed in Section 6.4 to obtain the filtered path of regime denoted by $Y_{\max|u}$, and then we store the values of U and $Y_{\max|u}$ as a pair. Given the terminal value of the price process, as long as U is obtained by step (i) of Algorithm 6.1, we can use the corresponding stored path of regime to obtain the path of volatility and the path of price process rather than generating them by steps (ii) and (iii) of Algorithm 6.1.

(iii) For the problem of sampling conditioned processes of model (6.1), we assume that the volatility follows a continuous-time Markov chain. We can also consider the same sampling problem for hidden Markov models, where the Markov chain is discrete-time, and hence the jump events can only occur at a pre-determined set of points.

Appendices

Appendix for Chapter 2

2.A The EM Algorithm

We want to compare the estimation results obtained by the DECF estimation method with a commonly used approach, such as the EM method. In the following, we will present a modified EM algorithm to estimate regime switching models whose distributions are characterized by their characteristic functions only.

Kim (1994) has proposed an EM algorithm that can find the maximum likelihood estimates of the parameters in models with unobserved (latent) variables. The EM algorithm alternates between the E-step and the M-step until the resulting log-likelihood converges. The E-step contains two sub-steps, filtering and smoothing, and it creates a function used in the expected log-likelihood function whose parameters are the current estimates. The M-step maximizes the expected log-likelihood to obtain a new set of estimated parameters, which are then used in the next E-step.

Define $\theta^{(0)} = (\underline{\xi}^{(0)}, P^{(0)}, \rho_k^{(0)}), \quad k = 1, ...K$, as the initial parameter vector for the estimation algorithm, where $\underline{\xi}^{(0)}$ is a set of model parameters, $P^{(0)}$ is the transition matrix of the regime switching model defined in (1.1), and $\rho_k^{(0)} = P(Y_1 = k)$ is the probability that the initial state of the model is k. We assume that the parameter vector $\theta^{(n)}$ is obtained in the M-step of the previous n^{th} iteration, and $w = \{w_j, j = 1, ..., N\}$ is the realization of the process $\mathcal{W} = \{W_j, j = 1, ..., N\}$.

Following Kim (1994), below we present the EM algorithm to estimate a regime switching model.

1. **E-Step (Filtering)**

For $j = 1, 2, ..., N$, we iterate equations

$$P(Y_j = k|w_j, \theta^{(n)}) = \frac{P(Y_j = k|w_{j-1}, \theta^{(n)})f(w_j|w_{j-1}, Y_j = k, \theta^{(n)})}{\sum_{k=1}^{K} P(Y_j = k|w_{j-1}, \theta^{(n)})f(w_j|w_{j-1}, Y_j = k, \theta^{(n)})}, \tag{1}$$

where $f(w_j|w_{j-1}, Y_j = k, \theta^{(n)})$ is the conditional density function of w_j given w_{j-1} at regime k, and

$$P(Y_{j+1} = k|w_j, \theta^{(n)}) = \sum_{l=1}^{K} p_{lk} P(Y_j = l|w_j, \theta^{(n)}), \tag{2}$$

until $P(Y_N = k|w_N, \theta^{(n)})$ is obtained with the starting point $P(Y_1 = k|w_0, \theta^{(n)}) = \rho_k^{(n)}$.

2. **E-Step (Smoothing)**

   For $j = N-1, N-2, ..., 1$, we iterate equations

$$P(Y_j = k|w_N, \theta^{(n)}) = \sum_{l=1}^{K} \frac{P(Y_j = k|w_j, \theta^{(n)})P(Y_{j+1} = l|w_N, \theta^{(n)})p_{kl}^{(n)}}{P(Y_{j+1} = l|w_j, \theta^{(n)})}. \tag{3}$$

3. **M-Step**

   Now, we can obtain $\underline{\xi}^{(n+1)}$ by maximizing the log-likelihood function

$$\log[L(\underline{\xi}^{(n+1)})] = \sum_{k=1}^{K} \sum_{j=1}^{N} P(Y_j = k|w_N, \theta^{(n)}) \log[f(w_j|w_{j-1}, Y_j = k, \underline{\xi}^{(n)})]. \tag{4}$$

Lastly,

$$\rho_k^{(n+1)} = P(Y_1 = k|w_N, \theta^{(n)}), \tag{5}$$

and

$$\begin{aligned} p_{kl}^{(n+1)} &= \frac{\sum_{j=2}^{N} P(Y_j = l, Y_{j-1} = k|w_N, \theta^{(n)})}{\sum_{j=2}^{N} P(Y_{j-1} = k|w_N, \theta^{(n)})} && (6) \\ &= \frac{\sum_{j=2}^{N} P(Y_j = l|w_N, \theta^{(n)}) \frac{p_{kl}^{(n)} P(Y_{j-1}=k|w_{j-1}, \theta^{(n)})}{P(Y_j=l|w_{j-1}, \theta^{(n)})}}{\sum_{j=2}^{N} P(Y_{j-1} = k|w_N, \theta^{(n)})}. && (7) \end{aligned}$$

Then we use $\theta^{(n+1)} = (\underline{\xi}^{(n+1)}, P^{(n+1)}, \rho_k^{(n+1)})$, $k = 1, ...K$, as the new parameter vector in the next iteration from the E-step.

For a model that is characterized by its characteristic functions only, the conditional density function $f(w_j|w_{j-1}, Y_j = k)$ in the above algorithm is usually unknown. Thus, we propose to use Fourier cosine series expansions to approximate the conditional density function.

For comparison purposes, we only consider Lévy processes here. Increments of a Lévy process are independent and stationary, so calculating the conditional density function $f(w_j|w_{j-1}, Y_j = k)$ is reduced to calculating the density function $f(w_j|Y_j = k)$.

To approximate $f(w_j|Y_j = k)$, we truncate the support of the density to an interval $[a, b]$, and use Fourier cosine series expansions with a finite number of terms L. Then,

$$\hat{f}(w_j|k) := \frac{A_0(k)}{2} + \sum_{u=1}^{L} A_u(k) \cdot \cos(u\pi \frac{w_j - a}{b - a}) \tag{8}$$

with

$$A_u(k) := \frac{2}{b-a} \int_a^b \hat{f}(x|k) \cos(u\pi \frac{x-a}{b-a}) dx, \quad u = 0, 1, \ldots, L. \tag{9}$$

The above cosine series coefficients can be approximated by using the characteristic function $\phi(\cdot; \underline{\xi}_k)$ of W at regime k as follows

$$A_u(k) \approx F_u(k) := \frac{2}{b-a} \Re \left\{ \phi(\frac{u\pi}{b-a}; \underline{\xi}_k) \cdot \exp(-i \frac{ua\pi}{b-a}) \right\}. \tag{10}$$

We only have an approximation not equality in (10), which is due to the truncation of the integration region in the definition of the characteristic function. Moreover, we need to normalize the approximation of the density function (8) so that its integral over the truncated range is 1. Then we can approximate the conditional density function $f(w_j|w_{j-1}, Y_j = k)$ by the following function:

$$\hat{f}(w_j|w_{j-1}, Y_j = k) := \frac{\hat{f}(w_j|k)}{\sum_{u=0}^{L} \hat{f}(a + u * \frac{b-a}{L}|k)}. \tag{11}$$

2.B Proof of Lemma 2.2.2.

Substituting (2.14) into formula (2.12), we obtain

$$
\bar{\psi}_k(r_i, r_l) = \begin{cases} \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3}^{(k-1)} p_{k_3,k_4} \\ \cdot \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4}),\ 2 \le k \le k^* & (12) \\ \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_1} p_{k_1,k_2} \pi_{k_3} p_{k_3,k_4} \\ \cdot \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4}), \ \ k \ge k^* + 1. & (13) \end{cases}
$$

Since $\pi_{k_3} p_{k_3,k_4} \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4})$ does not depend on k_1 and k_2, so $\bar{\psi}_k(r_i, r_l)$ can be rewritten as

$$
\bar{\psi}_k(r_i, r_l) = \begin{cases} \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3}^{(k-1)} p_{k_3,k_4} \\ \cdot \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4}),\ 2 \le k \le k^* & (14) \\ \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \pi_{k_1} p_{k_1,k_2} \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \\ \cdot \sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_3} p_{k_3,k_4} \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4}), \quad k \ge k^* + 1. & (15) \end{cases}
$$

By the definition of the error term, ϵ_{k,k^*} becomes

$$
\epsilon_{k,k^*} = \begin{cases} 0,\ 2 \le k \le k^* \\ |\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\sum_{k_3=1}^{K}\sum_{k_4=1}^{K} \pi_{k_1} p_{k_1,k_2} (p_{k_2,k_3}^{(k-1)} - \pi_{k_3}) p_{k_3,k_4} \\ \cdot \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4})|, \ \ k \ge k^* + 1. & (16) \end{cases}
$$

For all $\epsilon^* > 0$ and any $k^* \ge 1$, $\epsilon_{k,k^*} = 0 < \epsilon_*$ when $k \le k^*$. Thus, we need to prove that there exists k^* such that equation (16) is less than or equal to ϵ^* when $k \ge k^* + 1$.

The characteristic function ϕ is finite and bounded as $|\phi| \le 1$. In addition, the transition probabilities and the density of the stationary distribution are also finite and bounded by 1. Therefore,

$$
|\pi_{k_1} p_{k_1,k_2} p_{k_3,k_4} \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4})| \le C_{r_i, r_l}, \tag{17}
$$

where C_{r_i,r_l} is a positive and finite constant. Thus, we can rewrite equation (16) as

$$\epsilon_{k,k^*} \leq C_{r_i,r_l} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} \sum_{k_4=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|. \tag{18}$$

Since $|p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|$ does not depend on k_1 and k_4, this implies

$$\epsilon_{k,k^*} \leq C_{r_i,r_l} K^2 \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|, \tag{19}$$

or equivalently

$$\epsilon_{k,k^*} \leq C^* \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|, \quad \text{where} \quad C^* := C_{r_i,r_l} K^2. \tag{20}$$

From Perron-Frobenius Theorem, we know that $|p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|$ is a decreasing function of k and $\lim_{k\to\infty} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0$ for any k_2 and k_3. Because C_{r_i,r_l} and K are finite, so $C^* := C_{r_i,r_l} K^2$ is finite. Then, for any k_2 and k_3, $\lim_{k\to\infty} C^* |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0$, and hence $\lim_{k\to\infty} C^* \max_{\forall k_2,k_3} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0$.

Then we have

$$0 \leq \lim_{k\to\infty} C^* \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| \leq \lim_{k\to\infty} C^* \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} \max_{\forall k_2,k_3} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|,$$

and hence

$$0 \leq \lim_{k\to\infty} C^* \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| \leq \lim_{k\to\infty} C^* K^2 \max_{\forall k_2,k_3} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}|.$$

Because $\lim_{k\to\infty} C^* \max_{\forall k_2,k_3} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0$ and K^2 is finite, so $\lim_{k\to\infty} C^* K^2 \max_{\forall k_2,k_3} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0$. Therefore,

$$0 \leq \lim_{k\to\infty} C^* \sum_{k_2=1}^{K} \sum_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| \leq \lim_{k\to\infty} C^* K^2 \max_{\forall k_2,k_3} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0,$$

and by the Squeeze Theorem, $\lim\limits_{k\to\infty} C^* \sum\limits_{k_2=1}^{K} \sum\limits_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| = 0$.

This shows for any $\epsilon^* > 0$, we can find a k^* such that $\epsilon_{k,k^*} \leq C^* \sum\limits_{k_2=1}^{K} \sum\limits_{k_3=1}^{K} |p_{k_2,k_3}^{(k-1)} - \pi_{k_3}| \leq \epsilon^*$ for any $k \geq k^* + 1$. $\square$

2.C Proof of Lemma 2.2.3.

When $k^* = 1$, for any integer $k \geq 2$, equation (2.15) becomes

$$\bar{\psi}_k(r_i, r_l) = \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \pi_{k_1} p_{k_1,k_2} \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2}) \cdot \sum_{k_3=1}^{K} \sum_{k_4=1}^{K} \pi_{k_3} p_{k_3,k_4} \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4}). \tag{21}$$

Since by assumption, $\mathbf{W_{1,m}}$ and $\mathbf{W_{k+1,m}}$ are independent for $k \geq 2$, then (2.12) becomes:

$$\psi_k^*(r_i, r_l) \quad = \quad E[e^{(ir_i(W_1,W_2)^{'})}] \cdot E[e^{(ir_l(W_{k+1},W_{k+2})^{'})}]. \tag{22}$$

By conditioning, the right-hand side of (22) becomes

$$[\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \pi_{k_1} p_{k_1,k_2} \phi(r_{i1}; \underline{\xi}_{k_1}) \phi(r_{i2}; \underline{\xi}_{k_2})] \cdot [\sum_{k_3=1}^{K} \sum_{k_4=1}^{K} \pi_{k_3} p_{k_3,k_4} \phi(r_{l1}; \underline{\xi}_{k_3}) \phi(r_{l2}; \underline{\xi}_{k_4})]. \tag{23}$$

We can see that (23) and (2.15) are in the same form. Thus, for any $k \geq 2$, if we assume that $\mathbf{W_{1,m}}$ and $\mathbf{W_{k+1,m}}$ are independent, then (2.12) is the same as (2.15) with $k^* = 1$. $\square$

2.D Proof of Lemma 2.2.4.

When we condition (2.23) on Y_{j-2}, Y_{j-1} and Y_j, for the LHS:

$$\begin{aligned}
& P(W_{j-2} = w_{j-2}, W_j = w_j | W_{j-1} = w_{j-1}) \\
= \quad & \sum_{y_{j-2}=1}^{K} \sum_{y_{j-1}=1}^{K} \sum_{y_j=1}^{K} P(W_{j-2} = w_{j-2} | Y_{j-2} = y_{j-2}) P(W_j = w_j | Y_j = y_j) p_{y_{j-1},y_j} \cdot \\
& P(Y_{j-1} = y_{j-1} | W_{j-1} = w_{j-1}, Y_{j-2} = y_{j-2}) P(Y_{j-2} = y_{j-2} | W_{j-1} = w_{j-1}).
\end{aligned}$$

Similarly, for the RHS:

$$
\begin{aligned}
& P(W_{j-2}=w_{j-2}|W_{j-1}=w_{j-1})P(W_j=w_j|W_{j-1}=w_{j-1}) \\
= \; & \Big\{\sum_{y_{j-2}=1}^{K} P(W_{j-2}=w_j|W_{j-1}=w_{j-1},Y_{j-2}=y_{j-2})P(Y_{j-2}=y_{j-2}|W_{j-1}=w_{j-1})\Big\} \\
& \cdot\Big\{\sum_{y_{j-1}=1}^{K}\sum_{y_j=1}^{K} P(W_j=w_j|Y_j=y_j)p_{y_{j-1},y_j}P(Y_{j-1}=y_{j-1}|W_{j-1}=w_{j-1})\Big\} \\
= \; & \sum_{y_{j-2}=1}^{K}\sum_{y_{j-1}=1}^{K}\sum_{y_j=1}^{K} P(W_{j-2}=w_{j-2}|W_{j-1}=w_{j-1},Y_{j-2}=y_{j-2})\cdot \\
& P(Y_{j-2}=y_{j-2}|W_{j-1}=w_{j-1})P(W_j=w_j|Y_j=y_j)p_{y_{j-1},y_j}P(Y_{j-1}=y_{j-1}|W_{j-1}=w_{j-1}) \\
= \; & \sum_{y_{j-2}=1}^{K}\sum_{y_{j-1}=1}^{K}\sum_{y_j=1}^{K} P(W_{j-2}=w_{j-2}|Y_{j-2}=y_{j-2})P(W_j=w_j|Y_j=y_j)p_{y_{j-1},y_j}\cdot \\
& P(Y_{j-1}=y_{j-1}|W_{j-1}=w_{j-1})P(Y_{j-2}=y_{j-2}|W_{j-1}=w_{j-1}).
\end{aligned}
$$

Thus, based on our definition and assumption for the term $\overset{c.l.}{\sim}$, condition (2.23) holds if for any $j=3,...,N$,

$$P(Y_{j-1}=y_{j-1}|W_{j-1}=w_{j-1}) \overset{c.l.}{\sim} P(Y_{j-1}=y_{j-1}|W_{j-1}=w_{j-1},Y_{j-2}=y_{j-2}). \quad \square$$

2.E Proof of Lemma 2.2.5.

For the LHS of condition (2.24):

$$
\begin{aligned}
P(Y_{j-1}=y_{j-1}|W_{j-1}=w_{j-1}) &= \frac{P(W_{j-1}=w_{j-1},Y_{j-1}=y_{j-1})}{P(W_{j-1}=w_{j-1})} \\
&= \frac{P(W_{j-1}=w_{j-1}|Y_{j-1}=y_{j-1})P(Y_{j-1}=y_{j-1})}{\sum_{y_{j-1}=1}^{2} P(W_{j-1}=w_{j-1}|Y_{j-1}=y_{j-1})P(Y_{j-1}=y_{j-1})},
\end{aligned}
$$

or

$$P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}) = \frac{P(W_{j-1} = w_{j-1}|Y_{j-1} = y_{j-1})\pi_{y_{j-1}}}{\sum_{y_{j-1}=1}^{2} P(W_{j-1} = w_{j-1}|Y_{j-1} = y_{j-1})\pi_{y_{j-1}}}.$$

Because stationary distributions can be represented in terms of transition probabilities as

$$\pi_1 = \frac{p_{21}}{p_{12} + p_{21}} \quad \text{and} \quad \pi_2 = \frac{p_{12}}{p_{12} + p_{21}},$$

then we have

$$P(Y_{j-1} = 1|W_{j-1} = w_{j-1}) = \frac{P(W_{j-1} = w_{j-1}|Y_{j-1} = 1)p_{21}}{P(W_{j-1} = w_{j-1}|Y_{j-1} = 1)p_{21} + P(W_{j-1} = w_{j-1}|Y_{j-1} = 2)p_{12}},$$

and

$$P(Y_{j-1} = 2|W_{j-1} = w_{j-1}) = \frac{P(W_{j-1} = w_{j-1}|Y_{j-1} = 2)p_{12}}{P(W_{j-1} = w_{j-1}|Y_{j-1} = 1)p_{21} + P(W_{j-1} = w_{j-1}|Y_{j-1} = 2)p_{12}}.$$

Similarly, for the RHS of condition (2.24):

$$\begin{aligned}
& P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}, Y_{j-2} = y_{j-2}) \\
= & \frac{P(W_{j-1} = w_{j-1}, Y_{j-1} = y_{j-1}, Y_{j-2} = y_{j-2})}{P(W_{j-1} = w_{j-1}, Y_{j-2} = y_{j-2})} \\
= & \frac{P(W_{j-1} = w_{j-1}|Y_{j-1} = y_{j-1})P(Y_{j-1} = y_{j-1}|Y_{j-2} = y_{j-2})P(Y_{j-2} = y_{j-2})}{\sum_{y_{j-1}=1}^{2} P(W_{j-1} = w_{j-1}|Y_{j-1} = y_{j-1})P(Y_{j-1} = y_{j-1}|Y_{j-2} = y_{j-2})P(Y_{j-2} = y_{j-2})}, \\
= & \frac{P(W_{j-1} = w_{j-1}|Y_{j-1} = y_{j-1})p_{y_{j-2},y_{j-1}}\pi_{y_{j-2}}}{\sum_{y_{j-1}=1}^{2} P(W_{j-1} = w_{j-1}|Y_{j-1} = y_{j-1})p_{y_{j-2},y_{j-1}}\pi_{y_{j-2}}},
\end{aligned}$$

we can represent the distribution $P(Y_{j-1} = y_{j-1}|W_{j-1} = w_{j-1}, Y_{j-2} = y_{j-2})$, for any $y_{j-1}, y_{j-2} \in \{1, 2\}$, in terms of model densities and parameters.

Note that (2.24) holds if for any $j = 3, ..., N$, the following four conditions hold:

(i)

$$P(Y_{j-1}=1|W_{j-1}=w_{j-1}) \overset{c.l.}{\sim} P(Y_{j-1}=1|W_{j-1}=w_{j-1}, Y_{j-2}=1).$$

(ii)

$$P(Y_{j-1}=1|W_{j-1}=w_{j-1}) \overset{c.l.}{\sim} P(Y_{j-1}=1|W_{j-1}=w_{j-1}, Y_{j-2}=2).$$

(iii)

$$P(Y_{j-1}=2|W_{j-1}=w_{j-1}) \overset{c.l.}{\sim} P(Y_{j-1}=2|W_{j-1}=w_{j-1}, Y_{j-2}=1).$$

(iv)

$$P(Y_{j-1}=2|W_{j-1}=w_{j-1}) \overset{c.l.}{\sim} P(Y_{j-1}=2|W_{j-1}=w_{j-1}, Y_{j-2}=2).$$

By combining conditions (i) and (iii) and (ii) and (iv), we obtain the two conditions (2.25) and (2.26) respectively. Thus, the results follow. □

2.F Proof of Lemma 2.2.6.

To simplify conditions (2.25) and (2.26), we first divide the numerators and denominators of $d_j, j=1,...,6$, by their numerators respectively. In addition, these compared distributions are in the neighbourhood of the true value of the parameter, then we can obtain the following sufficient conditions to (2.25) and (2.26) respectively:

(i)

$$\frac{d_{W,2}p_{12}}{d_{W,1}p_{21}} \overset{c.l.}{\sim} \frac{d_{W,2}p_{12}}{d_{W,1}p_{11}} \overset{c.l.}{\sim} \frac{d_{W,2}p_{22}}{d_{W,1}p_{21}},$$

(ii)

$$\frac{d_{W,1}p_{21}}{d_{W,2}p_{12}} \overset{c.l.}{\sim} \frac{d_{W,1}p_{11}}{d_{W,2}p_{12}} \overset{c.l.}{\sim} \frac{d_{W,1}p_{21}}{d_{W,2}p_{22}}.$$

Since condition (ii) is equivalent to (i), then we only need to consider condition (i). In addition, we can get a sufficient condition for condition (i) as

$$\frac{p_{12}}{p_{21}} \overset{c.l.}{\sim} \frac{p_{12}}{p_{11}} \overset{c.l.}{\sim} \frac{p_{22}}{p_{21}}. \tag{24}$$

If $\frac{p_{12}}{p_{11}} \overset{c.l.}{\sim} \frac{p_{22}}{p_{21}}$, then we have $p_{11} \overset{c.l.}{\sim} p_{21}$ and hence $\frac{p_{12}}{p_{21}} \overset{c.l.}{\sim} \frac{p_{12}}{p_{11}}$. Then condition (24) can be simplified as

$$\frac{p_{12}}{p_{11}} \overset{c.l.}{\sim} \frac{p_{22}}{p_{21}}.$$

Similarly, condition (i) can be simplified as

$$\frac{d_{W,2}p_{12}}{d_{W,1}p_{11}} \overset{c.l.}{\sim} \frac{d_{W,2}p_{22}}{d_{W,1}p_{21}}.$$

Thus, the results follow. $\square$

Appendix for Chapter 3

3.A Proof of Theorem 3.2.2.

Let $\mathcal{X}$ be the centered process $\mathcal{W}$, which is also stationary. To prove the result stated in Theorem 3.2.2, we are going to use Theorem 3.2.1 to show that $\mathcal{X}$ is asymptotically normally distributed. For this, we check condition (i) in Theorem 3.2.1.

First, assume that $\delta = 1$:

The first part of condition (i) holds because of the model assumptions. For the second part of condition (i), we can prove it by using Theorem 1 in Mackay (2002). This result states that the mixing coefficients for stationary hidden Markov models satisfy $\alpha_l = O(l^{-\nu})$ for some $\nu > 2q+1$, where $q \in \mathbb{Z}^+$. Thus, we can assume that the coefficients satisfy $\alpha_l \leq cl^{-\nu}$ for some positive constant c, and we need to check $\sum_{l=1}^{\infty} \alpha_l^{\frac{1}{3}} \leq \sum_{l=1}^{\infty} (cl^{-\nu})^{\frac{1}{3}} < \infty$.

Let $b_l = (cl^{-\nu})^{\frac{1}{3}}, l = 1, 2, 3, ...$ Since c is a positive constant and $\nu > 2q+1, q \in \mathbb{Z}^+$, then $b_l > 0$ for all l and

$$\sum_{l=1}^{\infty} b_l = \sum_{l=1}^{\infty} (cl^{-\nu})^{\frac{1}{3}} = c^{\frac{1}{3}} \sum_{l=1}^{\infty} (\frac{1}{l})^{\frac{\nu}{3}}.$$

The sum $\sum_{l=1}^{\infty}(\frac{1}{l})^{\frac{\nu}{3}}$ is a Riemann zeta function, or Euler-Riemann zeta function, and it converges. Since $\dfrac{\nu}{3} > 1$, then $\sum_{l=1}^{\infty} b_l < \infty$. Thus, $\sum_{l=1}^{\infty} \alpha_l^{\frac{1}{3}} \leq \sum_{l=1}^{\infty}(cl^{-\nu})^{\frac{1}{3}} = \sum_{l=1}^{\infty} b_l < \infty$.

This result shows that condition (i) in Theorem 3.2.1 holds. In addition, if the variance σ_*^2 exists, that is, $0 < \sigma_*^2 < \infty$ then the CLT is applicable to a stationary hidden Markov model. Therefore, the CLT is applicable to the assumed models.

Lastly, we can follow the steps presented by Koutrouvelis and Kellermeier (1981) and use a standard differential argument, combined with the CLT, to show that the quadratic form of test statistic has an asymptotic χ^2 distribution with $2q - p$ $(p < 2q)$ degrees of freedom, where q is the number of the points at which a characteristic function is evaluated and p is the number of parameters need to be estimated of the testing model. We take $p = 0$ when model parameters are given. $\square$

Appendix for Chapter 5

5.A Characteristic Functions

Here we provide characteristic functions $\Psi_h(\cdot; \underline{\xi})$ of the variable $\mu(\underline{\xi}) + Z(\underline{\xi})$ in model (5.1), where $Z(\xi)$ corresponds to an increment of a particular Lévy process over time interval of length h and $\mu(\xi)$ satisfies (5.3). We also give formulas for $R_j = \Re(\Psi_h(\frac{\pi j}{l_X}))$ and $I_j = \Im(\Psi_h(\frac{\pi j}{l_X}))$, which can be used to obtain the coefficients in each regime for the expansion (5.22).

Black-Scholes Model. The characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ is given by

$$\Psi_h(z; \underline{\xi}) \quad = \quad e^{iz(r-\sigma^2/2)h - \frac{\sigma^2 z^2}{2}h}$$

with $\underline{\xi} := \sigma$, which gives us the following coefficients

$$R_j = \Re\big(\Psi_h(\frac{\pi j}{l_X}; \underline{\xi})\big) = e^{-\frac{\sigma^2 h}{2}(\frac{\pi j}{l_X})^2} \cos\Big(\frac{\pi j}{l_X}(r - \frac{\sigma^2}{2})h\Big),$$

$$I_j = \Im\big(\Psi_h(\frac{\pi j}{l_X}; \underline{\xi})\big) = e^{-\frac{\sigma^2 h}{2}(\frac{\pi j}{l_X})^2} \sin\Big(\frac{\pi j}{l_X}(r - \frac{\sigma^2}{2})h\Big).$$

Merton's Model. The characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ is given by

$$\Psi_h(z;\underline{\xi}) = e^{i\mu(\underline{\xi})zh - \frac{1}{2}z^2\sigma^2 h + \lambda h(e^{-\frac{z^2\sigma_J^2}{2} + iz\mu_J} - 1)},$$

where $\underline{\xi} = (\sigma, \lambda, \sigma_J, \mu_J)$ and

$$\mu(\underline{\xi}) = r - \frac{\sigma^2}{2} - \lambda(e^{\frac{\sigma_J^2}{2} + \mu_J} - 1).$$

Hence,

$$\begin{aligned} R_j &= \exp\{-\frac{1}{2}(\frac{\pi j}{l_X})^2\sigma^2 h + \lambda h(\cos(\mu_J \frac{\pi j}{l_X})e^{-\frac{1}{2}\sigma_J^2(\frac{\pi j}{l_X})^2} - 1)\} \\ &\qquad \cdot \cos\left(\mu(\underline{\xi})\frac{\pi j}{l_X}h + \sin(\mu_J \frac{\pi j}{l_X})\lambda h e^{-\frac{1}{2}\sigma_J^2(\frac{\pi j}{l_X})^2}\right), \end{aligned}$$

$$\begin{aligned} I_j &= \exp\{-\frac{1}{2}(\frac{\pi j}{l_X})^2\sigma^2 h + \lambda h(\cos(\mu_J \frac{\pi j}{l_X})e^{-\frac{1}{2}\sigma_J^2(\frac{\pi j}{l_X})^2} - 1)\} \\ &\qquad \cdot \sin\left(\mu(\underline{\xi})\frac{\pi j}{l_X}h + \sin(\mu_J \frac{\pi j}{l_X})\lambda h e^{-\frac{1}{2}\sigma_J^2(\frac{\pi j}{l_X})^2}\right). \end{aligned}$$

Kou's Model. The characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ is

$$\Psi_h(z;\underline{\xi}) = e^{i\mu(\xi)zh - \frac{1}{2}z^2\sigma^2 h + iz\lambda h[\frac{p}{\lambda_+ - iz} - \frac{1-p}{\lambda_- + iz}]},$$

where $\underline{\xi} = (\sigma, \lambda, \lambda_+, \lambda_-, p)$ and

$$\mu(\xi) = r - \frac{\sigma^2}{2} - \lambda(\frac{p}{\lambda_+ - 1} - \frac{1-p}{\lambda_- + 1}).$$

Hence,

$$\begin{aligned} R_j &= \exp\{-\frac{1}{2}(\frac{\pi j}{l_X})^2\sigma^2 h - \lambda h(\frac{\pi j}{l_X})^2(\frac{p}{\lambda_+^2 + (\frac{\pi j}{l_X})^2} + \frac{1-p}{\lambda_-^2 + (\frac{\pi j}{l_X})^2})\} \\ &\qquad \cdot \cos\left(\mu(\xi)\frac{\pi j}{l_X}h + \lambda\frac{\pi j}{l_X}h(\frac{p\lambda_+}{\lambda_+^2 + (\frac{\pi j}{l_X})^2} - \frac{(1-p)\lambda_-}{\lambda_-^2 + (\frac{\pi j}{l_X})^2})\right), \end{aligned}$$

$$
\begin{aligned}
I_j &= \exp\{-\frac{1}{2}(\frac{\pi j}{l_X})^2\sigma^2 h - \lambda h(\frac{\pi j}{l_X})^2(\frac{p}{\lambda_+^2 + (\frac{\pi j}{l_X})^2} + \frac{1-p}{\lambda_-^2 + (\frac{\pi j}{l_X})^2})\} \\
&\qquad \cdot \sin\left(\mu(\xi)\frac{\pi j}{l_X}h + \lambda\frac{\pi j}{l_X}h(\frac{p\lambda_+}{\lambda_+^2 + (\frac{\pi j}{l_X})^2} - \frac{(1-p)\lambda_-}{\lambda_-^2 + (\frac{\pi j}{l_X})^2})\right).
\end{aligned}
$$

Variance Gamma Process. The characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ is given by

$$
\Psi_h(z;\underline{\xi}) = e^{izh\mu(\underline{\xi}) - \frac{h}{\nu}\log(1 - i\theta\nu z + \frac{1}{2}z^2\sigma^2\nu)},
$$

with $\underline{\xi} := (\sigma, \theta, \nu)$ and

$$
\mu(\underline{\xi}) = \left[r + \frac{1}{\nu}\log(1 - \theta\nu - \frac{1}{2}\sigma^2\nu)\right].
$$

It can be rewritten as

$$
\Psi_h(z;\underline{\xi}) = \exp\left(izh\mu(\underline{\xi}) - \frac{h}{\nu}(\log R + i\alpha)\right) = R^{-h/\nu}\left(\cos(zh\mu(\underline{\xi}) - \frac{h}{\nu}\alpha) + i\sin(zh\mu(\underline{\xi}) - \frac{h}{\nu}\alpha)\right)
$$

with

$$
R \equiv R(z) = \sqrt{(1 + \frac{1}{2}z^2\sigma^2\nu)^2 + (\theta\nu z)^2} \quad \text{and} \quad \alpha \equiv \alpha(z) = \operatorname{Arg}(1 - i\theta\nu z + \frac{1}{2}z^2\sigma^2\nu).
$$

This leads to

$$
\begin{aligned}
R_j &= R(\frac{\pi j}{l_X})^{-h/\nu} \cdot \cos(\frac{\pi j}{l_X}h\mu(\underline{\xi}) - h\frac{\pi j}{\nu l_X}\alpha), \\
I_j &= R(\frac{\pi j}{l_X})^{-h/\nu} \cdot \sin(\frac{\pi j}{l_X}h\mu(\underline{\xi}) - h\frac{\pi j}{\nu l_X}\alpha).
\end{aligned}
$$

Normal Inverse Gaussian Process. The characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ is given by

$$
\Psi_h(z;\underline{\xi}) = e^{izh\mu(\underline{\xi}) + izh\mu_0 + \delta h\left[\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + iz)^2}\right]}
$$

with $\underline{\xi} = (\mu_0, \delta, \alpha, \beta)$ and

$$
\mu(\underline{\xi}) = r - \delta[\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + 1)^2}] - \mu_0.
$$

It can be rewritten as

$$\Psi_h(z;\underline{\xi}) = e^{izh\mu(\underline{\xi})+izh\mu_0+\delta h\left[\sqrt{\alpha^2-\beta^2}-\sqrt{\rho}e^{i\phi/2}\right]},$$

where

$$\rho \equiv \rho(z) = \sqrt{(\alpha^2-\beta^2+z^2)^2+(2\beta z)^2} \quad \text{and} \quad \phi \equiv \phi(z) = \mathrm{Arg}(\alpha^2-(\beta+iz)^2).$$

This leads to

$$R_j = e^{\delta h(\sqrt{\alpha^2-\beta^2}-\sqrt{\rho_j}\cdot\cos(\phi_j/2))} \cdot \cos\left(\frac{\pi j}{l_X} h(\mu(\underline{\xi})+\mu_0) - \delta h\sqrt{\rho_j}\cdot\sin(\frac{\phi_j}{2})\right)$$

$$I_j = e^{\delta h(\sqrt{\alpha^2-\beta^2}-\sqrt{\rho_j}\cdot\cos(\phi_j/2))} \cdot \sin\left(\frac{\pi j}{l_X} h(\mu(\underline{\xi})+\mu_0) - \delta h\sqrt{\rho_j}\cdot\sin(\frac{\phi_j}{2})\right),$$

where

$$\rho_j = \rho(\frac{\pi j}{l_X}) \quad \text{and} \quad \phi_j = \phi(\frac{\pi j}{l_X}).$$

Tempered Stable Process. The characteristic function of $\mu(\underline{\xi}) + Z(\underline{\xi})$ is

$$\Psi_h(z;\underline{\xi}) = e^{izh\mu(\underline{\xi})+h\Gamma(-\alpha_+)c_+[(\lambda_+-iz)^{\alpha_+}-\lambda_+^{\alpha_+}]+h\Gamma(-\alpha_-)c_-[(\lambda_-+iz)^{\alpha_-}-\lambda_-^{\alpha_-}]},$$

where $\underline{\xi} = (\alpha_-, \alpha_+, \lambda_-, \lambda_+)$, $\Gamma(\cdot)$ is the gamma function and

$$\mu(\underline{\xi}) = r - \Gamma(-\alpha_+)c_+[(\lambda_+-1)^{\alpha_+} - \lambda_+^{\alpha_+}] - \Gamma(-\alpha_-)c_-[(\lambda_-+1)^{\alpha_-} - \lambda_-^{\alpha_-}].$$

It can be rewritten as

$$\Psi_h(z;\underline{\xi}) = e^{h\Gamma(-\alpha_+)c_+[\rho_+^{\alpha_+}\cos(\phi_+\alpha_+)-\lambda_+^{\alpha_+}]+h\Gamma(-\alpha_-)c_-[\rho_-^{\alpha_-}\cos(\phi_-\alpha_-)-\lambda_-^{\alpha_-}]}$$
$$\cdot e^{ih[z\mu(\underline{\xi})+\Gamma(-\alpha_+)c_+\rho_+^{\alpha_+}\sin(\phi_+\alpha_+)+\Gamma(-\alpha_-)c_-\rho_-^{\alpha_-}\sin(\phi_-\alpha_-)]},$$

where

$$\rho_+ \equiv \rho_+(z) = \sqrt{\lambda_+^2+z^2}, \quad \rho_- \equiv \rho_-(z) = \sqrt{\lambda_-^2+z^2},$$

and

$$\phi_+ \equiv \phi_+(z) = \mathrm{Arg}(\lambda_+ - iz), \quad \phi_- \equiv \phi_-(z) = \mathrm{Arg}(\lambda_- + iz).$$

Hence,

$$R_j = e^A \cos(B(j)) \qquad \text{and} \qquad I_j = e^A \sin(B(j)),$$

where

$$A = h\left[\Gamma(-\alpha_+)c_+[\rho_+^{\alpha_+}\cos(\phi_+\alpha_+) - \lambda_+^{\alpha_+}] + \Gamma(-\alpha_-)c_-[\rho_-^{\alpha_-}\cos(\phi_-\alpha_-) - \lambda_-^{\alpha_-}]\right]$$

and

$$B(j) = h[\frac{\pi j}{l_X}\mu(\underline{\xi}) + \Gamma(-\alpha_+)c_+\rho_+^{\alpha_+}\sin(\phi_+\alpha_+) + \Gamma(-\alpha_-)c_-\rho_-^{\alpha_-}\sin(\phi_-\alpha_-)].$$

Appendix for Chapter 6

6.A The Inverse Transform Method

Suppose we want to sample random variates from a CDF, denoted by F. Denote the inverse function of F by F^{-1}. For a uniformly generated random variate z from the interval $[0,1]$, we can obtain a sample from the distribution F by taking the inverse as $x = F^{-1}(z)$. If the inverse function is unknown, then the problem is reduced to finding an approximation of F^{-1}. Chen et al. (2011) propose the inverse transform method to solve this problem as follows.

Given an $\epsilon > 0$, select an interval $[x_0, x_K]$ such that $\max(F(x_0), 1 - F(x_K)) < \epsilon$. Let $h := \frac{x_K - x_0}{K}$ for a positive integer K, and $x_k = x_0 + kh, 0 \le k \le K$. Compute the values $F_k \equiv F(x_k)$ and store (x_k, F_k) as a pair for each $0 \le k \le K$. For a uniformly generated random variate z from the interval $[0,1]$, use the binary search to find k, where $0 \le k < K$, such that $F_k \le z < F_{k+1}$. Then $F^{-1}(z)$ can be approximated by the following linear representation:

$$F^{-1}(z) \approx x_k + \frac{x_{k+1} - x_k}{F_{k+1} - F_k}(z - F_k).$$

When $0 < z < F_0$ or $F_K \le z < 1$, we can use the convention shown as follows to approximate $F^{-1}(z)$ if ϵ is given small, say 10^{-8}. Then we have $F^{-1}(z) \approx x_0$ if $0 < z < F_0$ and $F^{-1}(z) \approx x_K$ if $F_K \le z < 1$.

6.B Proof of Proposition 6.1.1.

The regime does not change when $N = 0$, therefore $T_1 = T$ under the assumption that the initial state is 1. Then we have,

$$F_{T_1|N}(T_1 \leq l|N = 0) = \begin{cases} 1, & \text{if } l \geq T, \\ 0, & \text{otherwise.} \end{cases}$$

Given N, let $\underline{X} = \{X_1, ..., X_N\}$ denote the event times. By the property of a Poisson process, we know that $f_{\underline{X}|N}$ follows a uniform distribution. Then we have $f_{\underline{X}|N} \propto 1$. Therefore, for any integer $N \geq 1$, we can obtain $F_{T_1|N}(T_1 \leq l|N = n)$ by integrating 1 over all possible domains of event times, and then normalize the integration.

Here we provide the proof of the results for N up to 10. Let $Z_N, N = 1, ..., 10$, be the normalizer for the integration. We can calculate Z_N by letting $F_{T_1|T}(T_1 \leq T|N = 1) = 1$.

$N = 1$

$$F_{T_1|N}(T_1 \leq l|N = 1) = Z_1 \int_0^l dX_1 = Z_1 l,$$

and hence $Z_1 = \frac{1}{T}$. Then we have

$$F_{T_1|N}(T_1 \leq l|N = 1) = \frac{l}{T} = \sum_{j=0}^{\lceil \frac{1}{2} \rceil - 1} \frac{1!}{(1-j)!j!} \frac{l^{1-j}(T-l)^j}{T^1}.$$

$N = 2$

In this case, $T_1 = X_1 + T - X_2$, then X_1 and X_2 satisfy the following conditions

$$0 \leq X_1 + T - X_2 \leq l \text{ and } 0 \leq X_1 \leq X_2 \leq T.$$

Then we have our integration domain:

$$X_1 \in [0, l + X_2 - T] \text{ and } X_2 \in [T - l, T].$$

Therefore,

$$F_{T_1|N}(T_1 \leq l|N = 2) = Z_2 \int_{T-l}^{T} \int_0^{l+X_2-T} dX_1 dX_2 = Z_2 \frac{l^2}{2},$$

and hence $Z_2 = \frac{2}{T^2}$. Then we have

$$F_{T_1|N}(T_1 \le l|N=2) = \frac{l^2}{T^2} = \sum_{j=0}^{\left\lceil \frac{2}{2} \right\rceil - 1} \frac{2!}{(2-j)!j!} \frac{l^{2-j}(T-l)^j}{T^2}.$$

$N = 3$

Similarly as above, use $0 \le X_1 \le ... \le X_N \le T$ to represent T_1 with the condition $T_1 \le l$, and we have two domains for integration:

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, l] \end{cases} \quad \text{and} \quad \begin{cases} X_1 \in [0, l + X_2 - X_3] \\ X_2 \in [X_3 - l, X_3] \\ X_3 \in [l, T] \end{cases} .$$

Then

$$F_{T_1|N}(T_1 \le l|N=3) = Z_3 \frac{-2l^3 + 3l^2T}{3!},$$

and hence $Z_3 = \frac{3!}{T^3}$. Then we have

$$F_{T_1|N}(T_1 \le l|N=3) = \frac{-2l^3 + 3l^2T}{T^3} = \sum_{j=0}^{\left\lceil \frac{3}{2} \right\rceil - 1} \frac{3!}{(3-j)!j!} \frac{l^{3-j}(T-l)^j}{T^3}.$$

$N = 4$

The domains for integration are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4 + l - T] \\ X_4 \in [T - l, T] \end{cases} \quad \text{and} \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [X_3 + T - X_4 - l, X_3] \\ X_3 \in [X_4 + l - T, X_4] \\ X_4 \in [T - l, T]. \end{cases}$$

With $Z_4 = \frac{4!}{T^4}$, we have

$$F_{T_1|N}(T_1 \le l|N=4) = \frac{-3l^4 + 4l^3T}{T^4} = \sum_{j=0}^{\left\lceil \frac{4}{2} \right\rceil - 1} \frac{4!}{(4-j)!j!} \frac{l^{4-j}(T-l)^j}{T^4}.$$

$N = 5$

The domains for integration are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4 + l - X_5] \\ X_4 \in [X_5 - l, X_5] \\ X_5 \in [l, T] \end{cases}, \qquad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, l] \end{cases}$$

and

$$\begin{cases} X_1 \in [0, l + X_4 - X_5 + X_2 - X_3] \\ X_2 \in [X_3 + X_5 - X_4 - l, X_3] \\ X_3 \in [l + X_4 - X_5, X_4] \\ X_4 \in [X_5 - l, X_5] \\ X_5 \in [l, T]. \end{cases}$$

With $Z_4 = \frac{5!}{T^5}$, we have

$$F_{T_1|N}(T_1 \leq l | N = 5) = \frac{6l^5 - 15l^4 + 10l^3T^2}{T^5} = \sum_{j=0}^{\left\lceil \frac{5}{2} \right\rceil - 1} \frac{5!}{(5-j)!j!} \frac{l^{5-j}(T-l)^j}{T^5}.$$

$N = 6$

The integration domains are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6 + l - T] \\ X_6 \in [T - l, T] \end{cases}, \qquad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4 - X_5 - T + X_6 + l] \\ X_4 \in [T + X_5 - X_6 - l, X_5] \\ X_5 \in [X_6 + l - T, X_6] \\ X_6 \in [T - l, T] \end{cases}$$

and

$$\begin{cases} X_1 \in [0, X_4 - X_5 - T + X_6 + l + X_2 - X_3] \\ X_2 \in [X_3 - l - X_6 + T + X_5 - X_4, X_3] \\ X_3 \in [l + X_6 - T + X_5 - X_4, X_4] \\ X_4 \in [T + X_5 - X_6 - l, X_5] \\ X_5 \in [X_6 + l - T, X_6] \\ X_6 \in [T - l, T] \end{cases},$$

then we have

$$F_{T_1|N}(T_1 \leq l|N = 6) = \frac{10l^6 - 24l^5T + 15l^4T^2}{T^6}.$$

Hereafter, for convenience, $X_i - X_j$ is denoted by $X_{i,j}$.

$N = 7$

The integration domains are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6] \\ X_6 \in [0, X_7] \\ X_7 \in [0, l] \end{cases}, \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, l + X_{6,7}] \\ X_6 \in [X_7 - l, X_7] \\ X_7 \in [l, T] \end{cases},$$

$$\begin{cases} X_1 \in [0, l + X_{6,7} + X_{4,5} + X_{2,3}] \\ X_2 \in [X_{3,4} + X_{5,6} + X_7 - l, X_3] \\ X_3 \in [X_{6,7} + X_{4,5} + l, X_4] \\ X_4 \in [X_{5,6} - l + X_7, X_5] \\ X_5 \in [l + X_{6,7}, X_6] \\ X_6 \in [X_7 - l, X_7] \\ X_7 \in [l, T] \end{cases} \quad \text{and} \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_{6,7} + X_{4,5} + l] \\ X_4 \in [X_{5,6} + X_7 - l, X_5] \\ X_5 \in [+X_{6,7}, X_6] \\ X_6 \in [X_7 - l, X_7] \\ X_7 \in [l, T] \end{cases},$$

then we have

$$F_{T_1|N}(T_1 \leq l|N = 7) = \frac{l^7 + 7l^6(T - l) + 21l^5(T - l)^2 + 35l^4(T - l)^3}{T^7}.$$

$N = 8$

The integration domains are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6] \\ X_6 \in [0, X_7] \\ X_7 \in [0, X_8 + l - T] \\ X_8 \in [T - l, T] \end{cases}, \qquad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_{6,7} + X_{4,5} + l + X_8 - T] \\ X_4 \in [X_{5,6} + X_{78} - l + T, X_5] \\ X_5 \in [l + X_{6,7} + X_8 - T, X_6] \\ X_6 \in [X_{7,8} - l + T, X_7] \\ X_7 \in [l + X_8 - T, T] \\ X_8 \in [T - l, T] \end{cases},$$

$$\begin{cases} X_1 \in [0, l + X_{6,7} + X_{4,5} + X_{2,3} + X_8 - T] \\ X_2 \in [X_{3,4} + X_{5,6} + X_{7,8} - l + T, X_3] \\ X_3 \in [X_{6,7} + X_{4,5} + l + X_8 - T, X_4] \\ X_4 \in [X_{5,6} - l + X_{7,8} + T, X_5] \\ X_5 \in [l + X_{6,7} + X_8 - T, X_6] \\ X_6 \in [X_{7,8} - l + T, X_7] \\ X_7 \in [l + X_8 - T, T] \\ X_8 \in [T - l, T] \end{cases} \text{and} \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, l + X_{6,7} + X_8 - T] \\ X_6 \in [T + X_{7,8} - l, X_7] \\ X_7 \in [l - T + X_8, X_8] \\ X_8 \in [T - l, T] \end{cases},$$

then we have

$$F_{T_1|N}(T_1 \leq l | N = 8) = \frac{l^8 + 8l^7(T - l) + 28l^6(T - l)^2 + 56l^5(T - l)^3}{T^8}.$$

$N = 9$

The integration domains are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6] \\ X_6 \in [0, X_7] \\ X_7 \in [0, X_8] \\ X_8 \in [0, X_9] \\ X_9 \in [0, l] \end{cases}, \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6] \\ X_6 \in [0, X_7] \\ X_7 \in [0, X_{8,9} + l] \\ X_8 \in [X_9 - l, X_9] \\ X_9 \in [l, T] \end{cases},$$

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, l + X_{6,7} + X_{8,9}] \\ X_6 \in [X_{7,8} + X_9 - l, X_7] \\ X_7 \in [l + X_{8,9}, X_8] \\ X_8 \in [X_9 - l, X_9] \\ X_9 \in [l, T] \end{cases}, \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_{4,5} + X_{6,7} + X_{8,9} + l] \\ X_4 \in [X_{5,6} + X_{7,8} + X_9 - l, X_5] \\ X_5 \in [l + X_{6,7} + X_{8,9}, X_6] \\ X_6 \in [X_{7,8} + X_9 - l, X_7] \\ X_7 \in [l + X_{8,9}, X_8] \\ X_8 \in [X_9 - l, X_9] \\ X_9 \in [l, T] \end{cases}$$

$$\text{and} \quad \begin{cases} X_1 \in [0, X_{2,3} + X_{4,5} + X_{6,7} + X_{8,9} + l] \\ X_2 \in [X_{3,4} + X_{5,6} + X_{7,8} + X_9 - l, X_3] \\ X_3 \in [X_{4,5} + X_{6,7} + X_{8,9} + l, X_4] \\ X_4 \in [X_{5,6} + X_{7,8} + X_9 - l, X_5] \\ X_5 \in [l + X_{6,7} + X_{8,9}, X_6] \\ X_6 \in [X_{7,8} + X_9 - l, X_7] \\ X_7 \in [l + X_{8,9}, X_8] \\ X_8 \in [X_9 - l, X_9] \\ X_9 \in [l, T] \end{cases},$$

then we have

$$F_{T_1|N}(T_1 \le l | N = 9) = \frac{1}{T^9}(l^9 + 9l^8(T-l) + 36l^7(T-l)^2 + 84l^6(T-l)^3 + 126l^5(T-l)^4).$$

$N = 10$

The integration domains are

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6] \\ X_6 \in [0, X_7] \\ X_7 \in [0, X_8] \\ X_8 \in [0, X_9] \\ X_9 \in [0, l - T + X_{10}] \\ X_{10} \in [T - l, T] \end{cases}, \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_6] \\ X_6 \in [0, X_7] \\ X_7 \in [0, X_{8,9} + X_{10} + l - T] \\ X_8 \in [X_{9,10} + T - l, X_9] \\ X_9 \in [l - T + X_{10}, X_{10}] \\ X_{10} \in [T - l, T] \end{cases},$$

$$\begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_4] \\ X_4 \in [0, X_5] \\ X_5 \in [0, X_{6,7} + X_{8,9} + X_{10} + l - T] \\ X_6 \in [X_{7,8} + X_{9,10} - l + T, X_7] \\ X_7 \in [X_{8,9} + X_{10} + l - T, X_8] \\ X_8 \in [X_{9,10} + T - l, X_9] \\ X_9 \in [l - T + X_{10}, X_{10}] \\ X_{10} \in [T - l, T] \end{cases}, \quad \begin{cases} X_1 \in [0, X_2] \\ X_2 \in [0, X_3] \\ X_3 \in [0, X_{4,5} + X_{6,7} + X_{8,9} + X_{10} + l - T] \\ X_4 \in [X_{5,6} + X_{7,8} + X_{9,10} + T - l, X_5] \\ X_5 \in [X_{6,7} + X_{8,9} + X_{10} + l - T, X_6] \\ X_6 \in [X_{7,8} + X_{9,10} - l + T, X_7] \\ X_7 \in [X_{8,9} + X_{10} + l - T, X_8] \\ X_8 \in [X_{9,10} + T - l, X_9] \\ X_9 \in [l - T + X_{10}, X_{10}] \\ X_{10} \in [T - l, T] \end{cases}$$

$$\text{and} \quad \begin{cases} X_1 \in [0, X_{2,3} + X_{4,5} + X_{6,7} + X_{8,9} + X_{10} + l - T] \\ X_2 \in [X_{3,4} + X_{5,6} + X_{7,8} + X_{9,10} - l + T, X_3] \\ X_3 \in [X_{4,5} + X_{6,7} + X_{8,9} + X_{10} + l - T, X_4] \\ X_4 \in [X_{5,6} + X_{7,8} + X_{9,10} + T - l, X_5] \\ X_5 \in [X_{6,7} + X_{8,9} + X_{10} + l - T, X_6] \\ X_6 \in [X_{7,8} + X_{9,10} - l + T, X_7] \\ X_7 \in [X_{8,9} + X_{10} + l - T, X_8] \\ X_8 \in [X_{9,10} + T - l, X_9] \\ X_9 \in [l - T + X_{10}, X_{10}] \\ X_{10} \in [T - l, T] \end{cases},$$

then we have

$$F_{T_1|N}(T_1 \leq l|N = 10) = \frac{1}{T^{10}}(l^{10}+10l^9(T-l)+45l^8(T-l)^2+120l^7(T-l)^3+210l^6(T-l)^4).$$

These five distributions also satisfy the equation

$$F_{T_1|N}(T_1 \leq l|N = n) = \sum_{j=0}^{\lceil \frac{n}{2} \rceil - 1} \binom{n}{j} (\frac{l}{T})^{n-j}(1 - \frac{l}{T})^j.$$

Therefore, the results follow. $\square$

6.C Proof of Proposition 6.1.2.

For any integer $n \geq 1$, we consider two cases, when n is odd and when n is even.

Given $U = u$ and $N = n$, where n is odd, we have

$$\begin{aligned} f_{\underline{X}|N,U} &\propto f_{\underline{X},N,U} \\ &= f_{U|\underline{X},N} f_{\underline{X}|N} f_N \\ &= f_{U|\underline{X},N} \cdot \frac{T^{-n}}{n!} \cdot \frac{\lambda_\Delta^n e^{-\lambda_\Delta}}{n!}, \text{ if } 0 < X_1 < X_2 ... < X_n < T, \end{aligned}$$

where the last equality follows from the properties of a Poisson process. Since U and N are given, then

$$f_{\underline{X}|N,U} \propto f_{U|\underline{X},N}.$$

Moreover, if $X_1, ..., X_n$ are given, then U is fixed. Therefore,

$$f_{U|\underline{X},N} = \begin{cases} 1, & \text{if } \sum_{i=1}^{\frac{n+1}{2}} (X_{2i-1} - X_{2i-2}) = \frac{u-\sigma_2^2 T}{\sigma_1^2-\sigma_2^2}, \\ 0, & \text{otherwise,} \end{cases}$$

and hence

$$f_{\underline{X}|N,U} \propto \begin{cases} 1, & \text{if } \sum_{i=1}^{\frac{n+1}{2}} (X_{2i-1} - X_{2i-2}) = \frac{u-\sigma_2^2 T}{\sigma_1^2-\sigma_2^2} \text{ and } 0 < X_1 ... < X_n < T, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, when n is even,

$$f_{\underline{X}|N,U} \propto \begin{cases} 1, & \text{if } \sum_{i=1}^{\frac{n}{2}} (X_{2i} - X_{2i-1}) = \frac{u-\sigma_2^2 T}{\sigma_1^2-\sigma_2^2} \text{ and } 0 < X_1 ... < X_n < T, \\ 0, & \text{otherwise.} \end{cases}$$

Since the distribution is proportional to a constant, so $f_{\underline{X}|N,U}$ follows a uniform distribution. □

6.D Proof of Lemma 6.1.1.

Let us focus on the first case. Note that T_1 and $N_{1,2}$ can be represented in terms of U and N respectively. Also, $W_1, ..., W_{N_{1,2}}$ can be represented in terms of the event times $X_1, ..., X_N$. By Bayes' theorem and the change of variables theorem, we have

$$f_{\underline{W}|T_1,N_{1,2}} = \frac{f_{\underline{W},T_1,N_{1,2}}}{f_{T_1,N_{1,2}}} = \frac{f_{\underline{X},U,N}}{f_{U,N}} \cdot H = f_{\underline{X}|U,N} \cdot H,$$

where H is a ratio of two Jacobians:

$$H = \frac{\left|\frac{\partial(\underline{W},T_1,N_{1,2})}{\partial(\underline{X},U,N)}\right|}{\left|\frac{\partial(T_1,N_{1,2})}{\partial(U,N)}\right|}.$$

Since H is a constant and $f_{\underline{X}|U,N}$ is a uniform distribution by Proposition 6.1.2, and hence $f_{\underline{W}|T_1,N_{1,2}}$ is uniform over the interval $[0, T_1]$ if $0 < W_1 < ... < W_{N_{1,2}} < T_1$. Similarly, we can prove that given T_2 and $N_{2,1}$, the joint distribution of $\underline{J} = \{J_1, ..., J_{N_{2,1}-1}\}$ is uniform over the interval $[0, T_2]$ if $0 < J_1 < ... < J_{N_{2,1}-1} < T_2$. Therefore, we can generate random variates from the distributions $f_{\underline{W}|T_1,N_{1,2}}$ and $f_{\underline{J}|T_2,N_{2,1}}$ by sampling uniformly from the interval $[0, T_1]$ and $[0, T_2]$ respectively. □

6.E Proof of the Continuity of the Likelihood Ratio in (6.25).

To prove the likelihood ratio in (6.25) is continuous, we need to prove that $dQ^*(s)$ and $dP^*(s)$ are continuous at every $s \in \mathcal{S}_L$. Let

$$f_{S,T_1}(s,t) := \frac{1}{s\sqrt{2\pi\bar{\sigma}(t)}} \exp(-\frac{(\ln s - \ln S_0 - rT_h + \frac{1}{2}\bar{\sigma}(t))^2}{2\bar{\sigma}(t)}).$$

Since a log-normal density function is continuous, so $f_{S,T_1}(s,t)$ is continuous at every $s \in \mathcal{S}_L$. Since f_{T_1} does not depend on $s \in \mathcal{S}_L$, this shows that the product $f_{S,T_1} \cdot f_{T_1}$ is also continuous at every $s \in \mathcal{S}_L$.

Next we prove that the numerator of (6.25), $g(s) := \int_0^{T_h} f_{S,T_1}(s,t) \cdot f_{T_1}(T_1 = t)dt$, is continuous at every $s \in \mathcal{S}_L$. For a given $s_0 \in \mathcal{S}_L$, define

$$t_s^* := \arg \sup_{t\in[0,T_h]} \left| f_{S,T_1}(s,t) - f_{S,T_1}(s_0,t) \right|, \ s \in \mathcal{S}_L.$$

Since f_{S,T_1} is continuous at any $s \in \mathcal{S}_L$, then for any number $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\left| s - s_0 \right| < \delta,$$

then

$$\left| f_{S,T_1}(s,t_s^*) - f_{S,T_1}(s_0,t_s^*) \right| < \epsilon.$$

Since f_{T_1} is the density function of T_1, then we have

$$\begin{aligned} \left| g(s) - g(s_0) \right| &\leq \int_0^{T_h} \left| f_{S,T_1}(s,t_s^*) - f_{S,T_1}(s_0,t_s^*) \right| f_{T_1}(t)dt \\ &< \int_0^{T_h} \epsilon f_{T_1}(t)dt = \epsilon \int_0^{T_h} f_{T_1}(t)dt \leq \epsilon \cdot 1 = \epsilon. \end{aligned}$$

This implies that function g is continuous at s_0. Since the above is valid for any $s_0 \in \mathcal{S}_L$, and hence $g(s)$ is continuous at every $s \in \mathcal{S}_L$.

Similarly, we can prove that the denominator of (6.25) is also continuous at every $s \in \mathcal{S}_L$. In addition, the denominator is an integral of a product of two density functions, and hence $dP^*(s) \neq 0$, for every $s \in \mathcal{S}_{\mathcal{L}}$. Thus the result follows. $\square$

6.F Derivation of the Sampling Log-normal Distribution is Section 6.1.4.

Based on the results in (6.2) and (6.3) and given $S_0 = s_0, S_T = s_T$ and the volatility path $\{v_t\}$ obtained in Section 6.1.3, we know that $f(S_T|S_t)$, $f(S_t|S_0)$ and $f(S_T|S_0)$ follow log-normal distributions. Denote

$$\mu_1^* := r(T-t) - \frac{1}{2}\int_t^T \sigma_s^2 ds, \quad \sigma_1^* = \sqrt{\int_t^T \sigma_s^2 ds},$$

$$\mu_2^* := rt - \frac{1}{2}\int_0^t \sigma_s^2 ds \quad \text{and} \quad \sigma_2^* = \sqrt{\int_0^t \sigma_s^2 ds}.$$

Then we have

$$f(S_T|S_t) \sim LN(\ln y + \mu_1^*, (\sigma_1^*)^2), \quad \text{and} \quad f(S_t|S_0) \sim LN(\ln x + \mu_2^*, (\sigma_2^*)^2).$$

Through a simple algebra, we have

$$f(S_t = s_t|S_0 = s_0, S_T = s_T) \propto f(S_T = s_T|S_t = s_t) f(S_t = s_t|S_0 = s_0) = C_{x,z} \cdot f^*(y|x,z),$$

where $C_{x,z}$ is a constant and $f^*(y|x,z) \sim LN(\mu^*, (\sigma^*)^2)$ with

$$\mu^* = \frac{(\sigma_1^*)^2(\ln s_0 + \mu_2) + (\sigma_2^*)^2(\ln s_T - \mu_1)}{(\sigma_1^*)^2 + (\sigma_2^*)^2} \quad \text{and} \quad \sigma^* = \frac{\sigma_1^* \sigma_2^*}{\sqrt{(\sigma_1^*)^2 + (\sigma_2^*)^2}}.$$

Then for any $t \in \{t_i, i = 1, ..., n\}$, we can use a log-normal random variate generator to sample S_t from the conditional distribution f^* as derived above.

Bibliography

[1] Abate, J. and Whitt, W. (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing systems*, 10(1-2):5–87.

[2] Altman, R. M. (2004). Assessing the goodness-of-fit of hidden Markov models. *Biometrics*, 60(2):444–450.

[3] Andersen, L. and Broadie, M. (2004). Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234.

[4] Bacinello, A. R., Biffis, E., and Millossovich, P. (2010). Regression-based algorithms for life insurance contracts with surrender guarantees. *Quantitative Finance*, 10(9):1077–1090.

[5] Bakshi, G. and Madan, D. (2000). Spanning and derivative-security valuation. *Journal of Financial Economics*, 55(2):205–238.

[6] Ballotta, L. (2010). Efficient pricing of ratchet equity-indexed annuities in a variance-gamma economy. *North American Actuarial Journal*, 14(3):355–368.

[7] Bang, D. (2012). Applications of periodic and quasiperiodic decompositions to options pricing. *The Journal of Computational Finance*, 24(1):3–31.

[8] Barndorff-Nielsen, O. E. (1997a). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13.

[9] Barndorff-Nielsen, O. E. (1997b). Processes of normal inverse Gaussian type. *Finance and stochastics*, 2(1):41–68.

[10] Bastani, A., Ahmadi, Z., and Damircheli, D. (2013). A radial basis collocation method for pricing American options under regime-switching jump-diffusion models. *Applied Numerical Mathematics*, 65:79–90.

[11] Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701.

[12] Bladt, M. and Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):395–410.

[13] Boyarchenko, S. and Levendorskii, S. (2009). American options in regime-switching models. *SIAM Journal on Control and Optimization*, 48(3):1353–1376.

[14] Boyd, J. (1989). *Fourier and Chebyshev spectral methods*. Springer-Verlag.

[15] Boyle, P. P., Kolkiewicz, A. W., and Tan, K. S. (2001). Valuation of the reset options embedded in some equity-linked insurance products. *North American Actuarial Journal*, 5(3):1–18.

[16] Boyle, P. P., Kolkiewicz, A. W., and Tan, K. S. (2013). Pricing Bermudan options using low-discrepancy mesh methods. *Quantitative Finance*, 13(6):841–860.

[17] Bradley, R. C. (1985). On the central limit question under absolute regularity. *The Annals of Probability*, pages 1314–1325.

[18] Bradley, R. C. et al. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2(2):107–144.

[19] Bremaud, P. (1999). *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media.

[20] Broadie, M. and Glasserman, P. (2004). A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, 7(4):35–72.

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

[22] Cai, N., Kou, S., Liu, Z., et al. (2014). A two-sided Laplace inversion algorithm with computable error bounds and its applications in financial engineering. *Advances in Applied Probability*, 46(3):766–789.

[23] Caramellino, L. and Zanette, A. (2011). Monte Carlo methods for pricing and hedging American options in high dimension. *Risk and Decision Analysis*, 2(4):207–220.

[24] Carr, P., Geman, H., Madan, D. B., and Yor, M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3):345–382.

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

[26] Carrasco, M., Chernov, M., Florens, J. P., and Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2):529–573.

[27] Carrasco, M., Chernov, M., Ghysels, E., and Florens, J.-P. (2002). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. *Available at SSRN 338961.*

[28] Carrasco, M. and Florens, J. P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16(06):797–834.

[29] Carriere, J. F. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: mathematics and Economics*, 19(1):19–30.

[30] Chen, S. X., Peng, L., Cindy, L. Y., et al. (2013). Parameter estimation and model testing for Markov processes via conditional characteristic functions. *Bernoulli*, 19(1):228–251.

[31] Chen, Z., Feng, L., and Lin, X. (2011). Inverse transform method for simulating Lévy processes and discrete Asian options pricing. In *Proceedings of the Winter Simulation Conference*, pages 444–456. Winter Simulation Conference.

[32] Cheney, E. and Kincaid, D. (2012). *Numerical mathematics and computing.* Cengage Learning.

[33] Cherubini, U., Della Lunga, G., Mulinacci, S., and Rossi, P. (2010). *Fourier transform methods in finance*, volume 524. John Wiley & Sons.

[34] Chiarella, C., El-Hassan, N., and Kucera, A. (1999). Evaluation of American option prices in a path integral framework using Fourier–Hermite series expansions. *Journal of Economic Dynamics and Control*, 23(9):1387–1424.

[35] Cinlar, E. (2013). *Introduction to stochastic processes.* Courier Dover Publications.

[36] Date, P., Mamon, R., and Tenyakov, A. (2013). Filtering and forecasting commodity futures prices under an HMM framework. *Energy Economics*, 40:1001–1013.

[37] Delyon, B. and Hu, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications*, 116(11):1660–1675.

[38] Detemple, J. (2010). *American-style derivatives: Valuation and computation*. CRC Press.

[39] Duan, J. C., Popova, I., and Ritchken, P. (2002). Option pricing under regime switching. *Quantitative Finance*, 2(116-132):209.

[40] Dufresne, D., Garrido, J., and Morales, M. (2009). Fourier inversion formulas in option pricing and insurance. *Methodology and Computing in Applied Probability*, 11(3):359–383.

[41] Durham, G. B. (2006). Monte Carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models. *Journal of Econometrics*, 133(1):273–305.

[42] Eberlein, E. (2014). Fourier-based valuation methods in mathematical finance. In *Quantitative energy finance*, pages 85–114. Springer.

[43] Elliott, R. J., Chan, L., and Siu, T. K. (2005). Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1(4):423–432.

[44] Epps, T. W. and Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, 70(3):723–726.

[45] Fang, F. and Oosterlee, C. W. (2008). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848.

[46] Fang, F. and Oosterlee, C. W. (2009). Pricing early-exercise and discrete barrier options by Fourier-cosine series expansions. *Numerische Mathematik*, 114(1):27–62.

[47] Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics*, 18(2):155–161.

[48] Feuerverger, A. and McDunnough, P. (1981). On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 20–27.

[49] Feuerverger, A. and McDunnough, P. (1984). On statistical transform methods and their efficiency. *Canadian Journal of Statistics*, 12(4):303–317.

[50] Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications. *The Annals of Statistics*, pages 88–97.

[51] Fuh, C. D., Ho, K. W. R., Hu, I., and Wang, R. H. (2012). Option pricing with Markov switching. *Journal of Data Science*, 10:483–509.

[52] Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

[53] Gelman, A. and Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185.

[54] Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199–213.

[55] Genest, C., Rémillard, B., et al. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de lInstitut Henri Poincaré: Probabilités et Statistiques*, 44(6):1096–1127.

[56] Gersho, A. and Gray, R. M. (1992). *Vector quantization and signal compression*. Springer Science & Business Media.

[57] Ghosh, S. and Ruymgaart, F. H. (1992). Applications of empirical characteristic functions in some multivariate problems. *Canadian Journal of Statistics*, 20(4):429–440.

[58] Golub, G. H. and Van Loan, C. F. (1996). Matrix computations. 1996. *Johns Hopkins University, Press, Baltimore, MD, USA*, pages 374–426.

[59] Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1):39–70.

[60] Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7(S1):S61–S82.

[61] Hardy, M. R. (2001). A regime-switching model of long-term stock returns. *North American Actuarial Journal*, 5(2):41–53.

[62] Hardy, M. R. (2002). Bayesian risk management for equity-linked insurance. *Scandinavian Actuarial Journal*, 2002(3):185–211.

[63] Hardy, M. R. (2003). *Investment guarantees: modeling and risk management for equity-linked life insurance*, volume 215. John Wiley & Sons.

[64] Harris, G. R. (1997). Regime switching vector autoregressions: a Bayesian Markov chain Monte Carlo approach. In *Proceedings of the 7th International AFIR Colloquium*, volume 1, pages 421–451.

[65] Hartman, B. M. and Groendyke, C. (2013). Model selection and averaging in financial risk management. *North American Actuarial Journal*, 17(3):216–228.

[66] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

[67] Heathcote, C. (1977). The integrated squared error estimation of parameters. *Biometrika*, 64(2):255–264.

[68] Hirsa, A. (2012). *Computational methods in finance*. CRC Press.

[69] Hobolth, A. and Stone, E. A. (2009). Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *The Annals of Applied Statistics*, 3(3):1204.

[70] Hörmann, W., Leydold, J., and Derflinger, G. (2013). *Automatic nonuniform random variate generation*. Springer Science & Business Media.

[71] Huang, Y., Forsyth, P. A., and Labahn, G. (2011). Methods for pricing American options under regime switching. *SIAM Journal on Scientific Computing*, 33(5):2144–2168.

[72] Jackson, K. R., Jaimungal, S., and Surkov, V. (2007). Option pricing with regime switching Lévy processes using Fourier space time stepping. In *Proc. 4th IASTED Intern. Conf. Financial Engin. Applic*, pages 92–97.

[73] Jackson, K. R., Jaimungal, S., and Surkov, V. (2008). Fourier space time-stepping for option pricing with Lévy models. *Journal of Computational Finance*, 12(2):1.

[74] Janczura, J. and Weron, R. (2013). Goodness-of-fit testing for the marginal distribution of regime-switching models with an application to electricity spot prices. *AStA Advances in Statistical Analysis*, 97(3):239–270.

[75] Janczura, J. and Weron, R. (2014). Inference for Markov regime-switching models of electricity spot prices. In *Quantitative Energy Finance*, pages 137–155. Springer.

[76] Jiang, G. J. and Knight, J. L. (2002). Estimation of continuous-time processes via the empirical characteristic function. *Journal of Business & Economic Statistics*, 20(2).

[77] Jiménez-Gamero, M. D., Alba-Fernández, V., Muñoz-García, J., and Chalco-Cano, Y. (2009). Goodness-of-fit tests based on empirical characteristic functions. *Computational Statistics & Data Analysis*, 53(12):3957–3971.

[78] Jin, X., Tan, H. H., and Sun, J. (2007). A state-space partitioning method for pricing high-dimensional American-style options. *Mathematical Finance*, 17(3):399–426.

[79] Kac, M. (1974). A stochastic model related to the telegrapher's equation. *Rocky Mountain J. Math*, 4:497–509.

[80] Këllezi, E. and Webber, N. (2004). Valuing Bermudan options when asset returns are Lévy processes. *Quantitative Finance*, 4(1):87–100.

[81] Kijima, M. and Wong, T. (2007). Pricing of ratchet equity-indexed annuities under stochastic interest rates. *Insurance: Mathematics and Economics*, 41(3):317–338.

[82] Kim, C. J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

[83] Knight, J. L. and Satchell, S. E. (1997). The cumulant generating function estimation method. *Econometric Theory*, 13(02):170–184.

[84] Knight, J. L., Satchell, S. E., et al. (1995). Estimation of stationary stochastic processes via the empirical characteristic function. Technical report, Faculty of Economics, University of Cambridge.

[85] Knight, J. L. and Yu, J. (2002). Empirical characteristic function in time series estimation. *Econometric Theory*, 18(03):691–721.

[86] Kolkiewicz, A. and Liu, Y. (2012). Semi-static hedging for GMWB in variable annuities. *North American Actuarial Journal*, 16(1):112–140.

[87] Kolkiewicz, A. W. (2016). Efficient static hedging of path-dependent options. Forthcoming in *International Journal of Theoretical and Applied Finance*.

[88] Konikov, M. and Madan, D. B. (2002). Option pricing using variance gamma Markov chains. *Review of Derivatives Research*, 5(1):81–115.

[89] Kotchoni, R. (2012). Applications of the characteristic function-based continuum GMM in finance. *Computational Statistics & Data Analysis*, 56(11):3599–3622.

[90] Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.

[91] Koutrouvelis, I. and Kellermeier, J. (1981). A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 173–176.

[92] Koutrouvelis, I. A. (1980). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, 67(1):238–240.

[93] Leucht, A. (2012). Characteristic function-based hypothesis tests under weak dependence. *Journal of Multivariate Analysis*, 108:67–89.

[94] Lin, X. S., Tan, K. S., and Yang, H. (2009). Pricing annuity guarantees under a regime-switching model. *North American Actuarial Journal*, 13(3):316–332.

[95] Longstaff, F. A. and Schwartz, E. S. (2001). Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies*, 14(1):113–147.

[96] López, O. and Ratanov, N. (2014). On the asymmetric telegraph processes. *Journal of Applied Probability*, 51(2):569–589.

[97] Lord, R., Fang, F., Bervoets, F., and Oosterlee, C. W. (2008). A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes. *SIAM Journal on Scientific Computing*, 30(4):1678–1705.

[98] Lu, Z. and Pong, T. K. (2011). Minimizing condition number via convex programming. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1193–1211.

[99] Luong, A. and Thompson, M. E. (1987). Minimum-distance methods based on quadratic distances for transforms. *Canadian Journal of Statistics*, 15(3):239–251.

[100] MacKay, R. J. (2002). Estimating the order of a hidden Markov model. *Canadian Journal of Statistics*, 30(4):573–589.

[101] Madan, D. B., Carr, P. P., and Chang, E. C. (1998). The variance gamma process and option pricing. *European Finance Review*, 2(1):79–105.

[102] Mamon, R. S. and Elliott, R. J. (2007). *Hidden Markov models in finance*, volume 4. Springer.

[103] Meng, X. L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, pages 831–860.

[104] Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144.

[105] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

[106] Naik, V. (1993). Option valuation and hedging strategies with jumps in the volatility of asset returns. *The Journal of Finance*, 48(5):1969–1984.

[107] Pagès, G., Pham, H., and Printems, J. (2004). Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, pages 253–297. Springer.

[108] Pagès, G. and Printems, J. (2008). Optimal quantization for finance: from random vectors to stochastic processes. *Handbook of Numerical Analysis*, 15:595–648.

[109] Printems, J. (2005). Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Applications MCMA*, 11(4):407–446.

[110] Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73(364):730–738.

[111] Rambharat, B. R. (2012). American option valuation with particle filters. In *Numerical Methods in Finance*, pages 51–82. Springer.

[112] Rémillard, B. (2011). Validity of the parametric bootstrap for goodness-of-fit testing in dynamic models. *Technical Report, SSRN Working Paper Series No 1966476.*

[113] Rémillard, B., Papageorgiou, N., and Soustra, F. (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis*, 110:30–42.

[114] Ruijter, M. and Oosterlee, C. (2012). Two-dimensional Fourier cosine series expansion method for pricing financial options. *SIAM Journal on Scientific Computing*, 34(5):B642–B671.

[115] Schmidt, P. (1982). An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica: Journal of the Econometric Society*, pages 501–516.

[116] Siu, T. K. (2005). Fair valuation of participating policies with surrender options and regime switching. *Insurance: Mathematics and Economics*, 37(3):533–552.

[117] Smith, R. D. (2007). An almost exact simulation method for the Heston model. *Journal of Computational Finance*, 11(1):115–125.

[118] Tankov, P. (2003). *Financial modelling with jump processes.* CRC press.

[119] Taufer, E., Leonenko, N., and Bee, M. (2011). Characteristic function estimation of Ornstein–Uhlenbeck-based stochastic volatility models. *Computational Statistics & Data Analysis*, 55(8):2525–2539.

[120] Tilley, J. A. (1993). Valuing American options in a path simulation model. *Transactions of the Society of Actuaries*, 45(83):104.

[121] Tone, C. (2010). A central limit theorem for multivariate strongly mixing random fields. *Probab. Math. Statist*, 30(2):215–222.

[122] Wei, X., Gaudenzi, M., and Zanette, A. (2013). Pricing ratchet equity-indexed annuities with early surrender risk in a CIR++ model. *North American Actuarial Journal*, 17(3):229–252.

[123] Wilmott, P., Dewynne, J., and Howison, S. (1993). *Option pricing: mathematical models and computation.* Oxford financial press.

[124] Yang, H. (2010). A numerical analysis of American options with regime switching. *Journal of Scientific Computing*, 44(1):69–91.

[125] Yin, G., Wang, J., Zhang, Q., and Liu, Y. (2006). Stochastic optimization algorithms for pricing American put options under regime-switching models. *Journal of Optimization Theory and Applications*, 131(1):37–52.

[126] Yu, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews*, 23(2):93–123.

[127] Yuen, F. L. and Yang, H. (2009). Option pricing in a jump-diffusion model with regime switching. *Astin Bulletin*, 39(02):515–539.

[128] Yuen, F. L. and Yang, H. (2010). Option pricing with regime switching by trinomial tree method. *Journal of Computational and Applied Mathematics*, 233(8):1821–1833.

[129] Zhu, J. (2009). *Applications of Fourier transform to smile modeling: Theory and implementation*. Springer.