In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

FINANCIAL FILTERING AND MODEL CALIBRATION

By

Ping Wu, B. Sc., M. Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

FINANCIAL FILTERING AND MODEL CALIBRATION

DOCTOR OF PHILOSOPHY (2003)
(Mathematics)

McMaster University
Hamilton, Ontario

TITLE: Financial Filtering and Model Calibration

AUTHOR: Ping Wu

B. Sc. (Southeast University, P. R. China)
M. Sc. (Southeast University, P. R. China)

SUPERVISOR: Professor Shui Feng

NUMBER OF PAGES: vi, 102

Abstract

The main topic of the thesis is the hidden Markov chain filtering for diffusions with jumps. The finite-dimensional filters are obtained for various statistics including the state of the hidden Markov chain. EM (Expectation Maximum) algorithm is then applied to the estimation of parameters of the hidden Markov chain. Based on some financial phenomena, we propose a model, called regime switching mean-reverting with jump model, for an asset price or interest rate. We apply the filtering methodology to this model and obtain the estimator for the mean reverting level. Monte Carlo simulation is performed in estimating the parameters of hidden Markov chain. The numerical methods for stochastic differential equations which are used in diffusion with jump model are discussed. Some fundamental results on stochastic calculus and some basic methods on generating random variables and sample paths are provided.

Acknowledgements

I would like to thank Dr. Shui Feng, my supervisor, for his guidance and supervision during this research.

I would also like to thank Dr. Tom Hurd for his many valuable suggestions and assistance when I study in Phimac Lab, and Dr. Roman Viveros, a member of my supervisory committee, for his constant encourage and interest in this work.

I also want to thank the following professors: Dr. Min-Oo, Dr. Guan, Dr. Balakrishnan, Dr. Hoppe, Dr. Alama, Dr. Gabardo and all members of the department of mathematics and statistics.

I express my thanks to MITACS, Dr. Shui Feng and McMaster University for providing the financial support for my graduate studies at McMaster University.

My thanks are due to Robert Elliott and Allanus Tsoi for their assistance.

I am indebted to my wife, Yu Bai , for her support and encourage that allowed me to finish my thesis in time. I wish to thank her for her unflagging support and confidence during the years of my study.

Finally, I wish to thank the following people: Jeff Boland, Tahir Choulli, Grasselli Matheus, Bogdan Traicu, Tao Li, Xuping Zhang, Joshua Fagbemi, and Symeon Scoufanic.

Hamilton, McMaster University — Ping Wu
September 28, 2002

Table of Contents

Chapter 1

Introduction

In 1960, R.E. Kalman published his famous paper describing a recursive solution to the discrete data linear filtering problem. Since that time, the filtering problem has been the subject of extensive research and application.

The filtering problem consists of estimating a stochastic process X_t, representing an unobserved signal, on the basis of the past and present observations $\{Y_s : 0 \leq s \leq t\}$ of a related measurement process Y. The information given by the measurement process up to time t is represented by the σ-algebra $\mathcal{Y}_t$ generated by $\{Y_s : 0 \leq s \leq t\}$. The solution of the filtering problem is the conditional density $P_{x_t|y_t}$ of the signal X_t given the observation Y_t. Such a solution in general takes its values in an infinite dimensional function space in an essential way, as proven in Chaleyat-Maurel and Michel [11]. As a consequence, in general the filtering cannot be implemented by an algorithm which updates only a finite number of parameters, this means that there can be no finite-memory computer implementation. The important exceptions are Kalman filter, Benes [7] filter, Elliott [30] finite state hidden Markov models and Kouritzin [48], [49] convolutional approach.

We explain the idea of filtering by Kalman filtering. We consider a discrete linear example:

$$X_{n+1} = A_n X_n + W_n \qquad n \geq 0,$$
$$Y_n = H_n X_n + V_n \qquad n \geq 0,$$

where $\{W_n\}$, $\{V_n\}$ are i.i.d. $N(0,1)$ uncorrelated sequences. We also assume that X_0 is normally distributed and independent of $\{W_n\}$, $\{V_n\}$. We want to get the optimal estimate (linear, unbiased, minimum variance) $\hat{X}_n$ using $\{Y_0, Y_1, \cdots, Y_n\}$. Since $\hat{X}_n$ is determined using all data $\{Y_0, Y_1, \cdots, Y_n\}$, the process is not really applicable to real-time problems for large values of n, since the need for storage of the data grows with time. However under the above assumption, we get a recursive algorithm such that $\hat{X}_n = f(\hat{X}_{n-1}, \xi_n)$, where ξ_n represents new information about X, when we obtain observation Y_n, the new information means the part of Y_n that is uncorrelated with $Y_1, Y_2, \cdots, Y_{n-1}$. In Kalman filtering, we obtain this "new information" by Hilbert Space Projection. Until $n-1$, we obtain the optimal estimate of X by using the data $\{Y_0, Y_1, \cdots, Y_{n-1}\}$, at n, and only use the incoming bit of the data information (new information) to correct the estimate of X so that no large storage of the data nor repetition of calculation is necessary.

In the sixties, several authors such as Bucy [10], Kushner [52], Wonham [76], Duncan [27], Mortensen [61], Zakai [77] generalized the results of Kalman filter from linear system theory to non-linear dynamics. This is an essentially more difficult problem, being in general infinite-dimensional, but equations describing the evolution of conditional distributions were obtained. In the seventies and the eighties, Fujisaki, Kallianpur and Kunita [39], Kallianpur [45], Di Masi and Runggaldier [57],

Pardox [63] and Davis [20] also had theoretical work in this area.

It is well-known that solving the non-linear filtering equation is difficult. In 1993, Elliott [30] utilized a finite state continuous time Markov chain instead of a general Markov process state equation in the filtering problem, namely using finite state continuous time Markov chain instead of the stochastic differential equation. The replacement of the actual Markov signal model with a possibly-infinite-dimensional signal had been previously considered by the likes of Kushner [53] and Di Masi and Runggaldier [56]. However, the finite-dimensional assumption facilitates simple calculation.

Recently, there have been many works devoted to applying various particle methods to construct approximate solutions to the celebrated Duncan-Mortensen-Zakai equations. Among them, we would like to mention the improved refining interacting particle filter by Del Moral, Kouritzin and Miclo [60] and the refining branching particle filter by Ballantyne, Chan and Kouritzin [3]. Kouritzin, Long and Sun [50] consider stochastic particle Markov chain approximation to the Duncan-Mortensen-Zakai equations for nonlinear filtering problems on regular, bounded domains. These works are part of the most recent progress in the theory of filtering.

Filtering techniques have been applied in areas as diverse as mathematical finance, target detection and tracking, communication networks, pollution tracking, weather prediction, traffic management, and search and rescue, and many others. Many papers and textbooks have been written on this subject since its inception in 1960. For

the most recent progress and development, we recommend the web site in [28].

Filtering problem is a huge field. In the following we will focus on the review of the filtering methodology used in the area of finance. These include bond market, commodity futures market, energy futures market, stock and option market, credit risk, and minimizing hedging strategy.

In the last ten years, filtering techniques have widely been applied to the estimation of increasingly complex financial models. There are four types of filtering techniques used in financial mathematics.

1. Kalman filtering and Extended Kalman filtering.
2. Non-linear filtering
3. Hidden Markov chain filtering
4. Quasi-maximum-likelihood approach combined with Kalman filtering.

1.1 Kalman Filtering and Extended Kalman filtering

The linear filtering problem (Kalman filter) is relative to the following model (limited to 1-dimensional, as the multi-dimensional version is a direct extension):

$$dX_t = F(t)X_t dt + C(t)dW_t; \quad F(t), C(t) \in \mathbb{R}, \tag{1.1.1}$$

$$dY_t = G(t)X_t dt + D(t)dV_t; \quad G(t), D(t) \in \mathbb{R}. \tag{1.1.2}$$

We assume that $F(t)$, $G(t)$, $C(t)$, $D(t)$ are bounded on bounded intervals. We also assume that X_0 is normally distributed and is independent of $\{W_t\}$, $\{V_t\}$. Finally we

assume that $D(t)$ is bounded away from 0 on bounded intervals.

Usually, an observed equation (1.1.2) is called the measurement system; this system represents the relationship between the observations $\{Y_t\}$ and the state variables $\{X_t\}$. The second , unobserved system of equation (1.1.1) is termed the transition system; this system describes the dynamic of the state variables $\{X_t\}$ as they were formulated in the model. Together, the measurement and transition equations represent what is called the state space form of the model. The Kalman filtering uses this state-space formulation to recursively make inference about the unobserved values of the state variables (transition system) by conditioning on the observed data (measurement system). As a final step, these recursive inferences are constructed and maximize a log-likelihood function to find the optimal parameter set.

The application of Kalman filtering methods in the estimation of term structure models has been investigated by many authors. Pennacchi [64] was the first to use this approach in financial econometrics. Exponential-affine models are considered in Chen and Scott [12], Chen and Scott [13], Babbs and Nowman [2], and Duan and Simonato [26]. This approach is very useful in situations where the underlying state variables are not observable. According to these models, bond price is a function of the unobserved instantaneous spot interest rate and the model parameters. The measurement system represents the relationship between zero-coupon rates and instantaneous spot interest rates. The transition system describes that the instantaneous spot interest rate follows the exponential-affine term structure model. In this

setting, the Kalman filter is applied to obtain inference about the instantaneous interest rate by conditioning on the observed market zero-coupon rates. The parameters are estimated by maximizing a log-likelihood function.

Schwartz [70] applies the Kalman filtering methodology to estimate the spot price and the parameter of three models (one, two, three-factor) for commercial commodities. One of the difficulties in the empirical implementation of commodity price models is that the spot price for some commodities is hard to obtain. However futures contracts are widely traded in several exchanges and their prices are more easily observed. The futures contract closet to maturity is used as a proxy for the spot price. The measurement equation is obtained by adding uncorrelated disturbances with mean zero to take into account bid-ask spreads, errors of the observations in the observed futures prices. The unobservable state variables (spot price) satisfy the transition equation which is generated by a Markov process. When the disturbances and the initial state vector are normally distributed, the Kalman filtering enables the likelihood function to be calculated and the spot price to be estimated.

Manoliu and Tompaidis [55] offer a general multi-factor model designed to account for the stochastic behavior of futures prices in energy market. They define energy futures prices in term of a spot price, not directly observable, driven by several stochastic factors. The spot price follows a generalized Ornstein-Uhlenbeck stochastic process. This formulation is well suited to the application of Kalman filtering techniques. They perform an empirical study of a one and a two-factor model for energy futures for natural gas based on the application of Kalman filtering techniques together with

maximum likelihood estimation methods.

The extended Kalman filtering is applied in credit risk by Cumby and Evans [17], and Claessens and Pennacchi [15]. They obtain the estimate of credit quality. The two approaches differ only in the update step. They treat credit quality as an unobserved variable that follows a specified stochastic process. The bond values are nonlinear functions of the unobservable variable. They take the average of the bond bid and ask prices as the observation data with measurement noise. Thus, a modification of the Kalman filter that considers functional nonlinearities, called the extended Kalman filter, is needed to obtain (approximate) maximum likelihood estimates of the model parameters. The extended Kalman filter uses the same recursive computational techniques as the Kalman filter, but linearizes the nonlinear function around the conditional mean of the state variables using a Taylor series expansion.

Lund [54] considers a nonlinear relationship between the observed data and the unobserved state variables. The main examples involve prices of coupon bond, and nonlinear term structure models such as exponential affine models. They utilize the extended Kalman filtering to estimate these models. In a Monte Carlo study they investigate the finite sample properties of the quasi maximum likelihood estimator for two term structure models. Schnatter [38] works on the same models with [54]. Their approach is to approximate the true update density by a Gaussian density with the same mean and variance as the exact update density. The mean and variance of the exact update density are computed by numerical integration. Evidence reported shows that this method is very effective to estimate a discrete-time log-normal

stochastic volatility model.

1.2 Non-linear Filtering

Volatility is probably the most important parameter in any financial model. Elliott, Hoekand and Valencia [33] consider the price process has dynamic $S_t = S_{t-1}\exp(\mu - \frac{1}{2}{\sigma_t}^2 + \sigma_t w_t^1)$ but the volatility follows the logarithmic mean-reverting process $\log \sigma_t = a + b\log \sigma_{t-1} + \theta w_t^2$. They use nonlinear filtering techniques to estimate volatility and the EM algorithm to calibrate the model parameters.

Tsoi, Yang and Yeung [74] assumed the risk-free interest rate follows a jump process and the stock price process depends on some underlying real variables. One example of the underlying is the market interest rate which is unobservable. In their paper, Cox-Ingersoll-Ross [16] model is used to model the market interest rate. They use counting process filtering technique to estimate the market interest rate from the information of the interest rate. Then they use the estimated value of market interest rate in their option price calculation.

Zeng [78] proposed a general micro-movement model that describes the transactional price behavior. The model is formulated as a filtering problem with counting process observations. In his paper, the filtering equations are derived. A theorem on the convergence of conditional expectation of the model is proven. A consistent recursive algorithm is constructed via the Markov chain approximation method.

Frey and Runggaldier [37] considered a market with a risky and a non-risky asset, and the price of the risky asset follows a stochastic volatility model. The paper describes an approach to determine a hedging strategy for an agent who has only incomplete price observations. Clearly, any reasonable strategy for this agent has to depend in some way on the unobservable latent state process. They obtain the risk minimizing hedging strategy under partial information by "projecting" the full information strategy onto the subfiltration. The subfiltration describes the available partial information that comes from observing the prices at the discrete random times where a trade occurs. To actually compute the projections onto the subfiltration, an important tool is the conditional distribution of the latent state process, given the available price observations. This leads to a filtering problem with counting process observations. It is an example of how filtering theory is applied in risk-minimizing hedging strategies. Such examples are rare in the literature.

1.3 Hidden Markov Chain Filtering

Hidden Markov models are playing a growing role in the discussion of stochastic phenomena. Filtering problems related to a finite state Markov chain are discussed in Elliott([30], [31]). In Elliott [30], they consider the process X_t is not observed directly, rather there is a (scalar) observation process given by

$$y_t = \int_0^t g(X_r)dr + W_t, \tag{1.3.1}$$

here $W = \{W_t : 0 \leq t\}$ is a standard Brownian motion which is independent of X_t. X_t, $t \geq 0$ be a finite state, continuous time, homogeneous Markov chain with

Q-matrix A. Therefore X_t has the semimartingale representation:

$$X_t = X_0 + \int_0^t AX_r dr + M_t, \tag{1.3.2}$$

where M_t is a $\mathcal{F}_t$-martingale. They take equation (1.3.1) as measurement system and equation (1.3.2) as transition system and obtain the finite dimensional filters. Elliott [31] is a discrete case of Elliott [30]. Recently, these results are used in financial mathematics, we describe the papers relative to this topic below.

Elliott and Rishel [36] assumed a risky asset whose price at time t is supposed to satisfy an equation of the form $dS_t = S_t(\rho_t dt + \sigma dW_t)$, where σ is constant. Further they assumed that the implicit interest rate ρ_t behaves like a continuous time finite state Markov chain. A natural process to take as the observation process is the (natural) logarithm of the price S_t, the results of Elliott [30] are applied to determine the implicit interest rate of the asset ρ_t and other parameters which describe its behavior.

A hidden Markov chain model with mean reverting characteristics is considered as a model for financial time series, particularly interest rates in Elliott, Fischer and Platen [32]. They consider a two factor mean reverting model, $dL_t = \gamma(\bar{L}_t - L_t)dt + \xi dW_t$, where γ, ξ are constants and the unobserved mean reversion level $\bar{L}_t$ changes according to a continuous time finite state Markov chain. In this setting they apply hidden Markov filtering and derive a finite dimensional filter for the unobservable state of Markov chain based on observations of the mean reverting diffusion process. A number of auxiliary filters are obtained that enable the parameters of the model to be estimated by EM algorithm. This is the first paper to consider a model for the short rate that is mean reverting to a stochastic level determined by

a Markov chain.

Elliott, Hunter and Jamieson [34] uses the filtering techniques in discrete time model. They assumed that a price process in discrete time evolves according to the dynamics $S_{n+1} = S_n \exp(\rho_n + \sigma_n w_n)$, where $\{w_n\}$, $n = 0, 1, 2, \cdots$, is a sequence of i.i.d $N(0, 1)$ random variables, and ρ_n and σ_n behave like a finite state Markov chain. Filtering and parameter estimation techniques in Elliott [31] are then applied to obtain the recursive estimates of ρ_n and σ_n. Further, all parameters in the model can be estimated. Unfortunately this technique can not be used in continuous time to estimate the volatility. This is because the probability measures corresponding to diffusions with different diffusion coefficients (volatility) maybe come singular with each other. However in discrete time, this method will work.

1.4 Quasi-maximum-likelihood Approach Combined with Kalman Filtering

This method aries from the stochastic volatility (SV) model, the signal satisfies the AR (autoregressive), ARMA (autoregressive moving average) or GARCH (generalized autoregressive conditional heteroscedastic) model. A key feature of the basic SV models in (1.4.1) below is that it can be transformed into a linear model by taking the logarithm of the squares of observations.

A simple asymptotically stationary SV model is given by

$$y_t = \varepsilon_t \exp(\frac{h_t}{2}), \tag{1.4.1}$$

$$h_t = \gamma + \phi h_{t-1} + \eta_t, \tag{1.4.2}$$

where y_t is the mean corrected return on holding the asset at time t, h_t is log volatility at time t, and ε_t is i.i.d. N(0,1) and η_t is i.i.d. N(0,σ^2). ε_t and η_t are uncorrelated. From (1.4.2), we have:

$$\log y_t^2 = \log \varepsilon_t^2 + h_t. \tag{1.4.3}$$

It is obvious that ε_t^2 is χ^2-distribution. The mean and variance of $\log \varepsilon_t^2$ are known to be -1.27 and $\pi^2/2 = 4.93$, respectively in Abramowitz and Stegun [1]. We rewrite the equation (1.4.3) into:

$$\log y_t^2 = -1.27 + h_t + \xi_t, \tag{1.4.4}$$

where $\xi_t = \log \varepsilon_t^2 + 1.27$ and $\text{var}(\xi_t) = \pi^2/2$. By treating ξ_t as if it were i.i.d. N(0, $\pi^2/2$), Harvey, Ruig and Shephard [43] considered equation (1.4.4) as measurement system and equation (1.4.2) as transition system, they employed Kalman filtering to estimate the unobservable log volatility, h_t, and used the quasi-likelihood function to perform parameter estimation. This quasi-likelihood estimator is suboptimal as ξ_t is poorly approximated by the normal distribution. (in this context see Ruig and Shephard [43], Kim and Shephard [46]).

Although Black-Scholes model has been widely used to study financial derivatives and the return of assets, many empirical investigations have suggested that the marginal distribution of the underlying assets has a higher peak and two heavier tails

than those of the normal distribution. Many authors, for example, Harrison and Pliska [42], and Harrison and Kreps [41], have extended the theory of asset price following Merton's [59] work, various models have arisen in financial mathematics such as stochastic volatility model and mean-reverting with jump model. Many authors have utilized the mean-reverting with jump model in interest rate and energy market (Das [18], Das [19], Dias [25], Deng [24] and Kou [47]). However, recent study shows that while the mean reverting level changes have an important effect on prices process (Hansen and Poulsen [40], Schlogl and Sommer [69]), the mean-reverting with jump model cannot explain more complex financial phenomena. We propose a type of model called regime-switching with jump model. In order that the model can be computationally tractable we choose the Poisson jump in our model, and the mean reverting level changes are governed by the continuous time finite state Markov chain. Based on this motivation, we first derive the hidden Markov chain filter for general diffusion with jumps in Chapter 4 , and then apply the result to the regime-switching model with jumps in Chapter 5.

A more detailed outline of this thesis is as follows:

In Chapter 2, we give a short presentation of stochastic calculus and its basic theory such as the generalized Itó formula, Girsanov's theorem, and Doléans-Dade's exponential formula which will be used in the following Chapter. Some basic concepts and notation are also defined as our preliminary settings, although they can be found in some classic books. Some examples are provided to illustrate the use of these formula and theorems. In the last section, we apply nonlinear filtering to estimate

volatility. It is a simple extended version of Elliott, Hoek and Valencia [33].

In Chapter 3, we introduce some basic random generator methods which include generators of Bernoulli distribution, exponential distribution, Poisson distribution, and others. We also present continuous time finite state Markov process simulation. All of them will be used in Chapter 5 for simulation study. We write several computer programs for several models including Brownian bridge process, and generalized Bessel process modulated by the Markov chain. We also extend the Milstein numerical method for diffusion model to diffusion with jumps model. In addition, some numerical results are compared.

In Chapter 4, we study the hidden Markov chain filtering for diffusion with jump model. We obtain recursive filtering for the latent states of the Markov chain, the number of jumps from one state to another, and the occupation time of the Markov chain in any state. The filtering equations are finite-dimensional and closed forms. We extend the EM algorithm to diffusion with jumps, and obtain the estimators of the parameters.

In Chapter 5, we consider a type of mean reverting model with jumps, where the mean reverting level changes according to a continuous time finite state Markov chain. This type of models can be used for modelling an asset price and an interest rate. The filtering techniques developed in Chapter 4 are applied to this model. We derive a finite dimensional filtering for the unobservable state of the Markov chain based on observations of the mean reverting diffusion with jumps process. Various

auxiliary filters are developed that allow us to estimate the parameters of the Markov chain. Simulation is done for a concrete example.

Chapter 2

Basic Theory

In this chapter we will review some fundamental results on martingales and semi-martingales and present some useful tools in stochastic calculus. These will include the generalized Itó formula, Girsanov's theorem and Doléans-Dade's exponential formula. Some basic definitions and notation are also provided as our preliminary setting (for more details see Protter [65], Oksendal [62], Elliott [29], and Brémaud [8]). We will also give some interesting examples arising in financial derivatives such as diffusion with jump model and stochastic volatility model.

2.1 Basic Definitions and Notation

We assume as given a complete probability space $(\Omega, \mathcal{F}, P)$. In addition we are given a filtration $(\mathcal{F}_t)_{0\leq t\leq\infty}$. By a filtration we mean a family of σ-algebras $(\mathcal{F}_t)_{0\leq t\leq\infty}$ that is increasing: $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$.

Definition 2.1.1. A filtered complete probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{0\leq t\leq\infty})$ is said to satisfy the usual hypotheses if
(i) $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$;
(ii) $\mathcal{F}_t = \cap_{u>t}\mathcal{F}_u$, all t, $0 \leq t < \infty$; that is, the filtration $(\mathcal{F}_t)_{0\leq t\leq\infty}$ is right continuous.

We always assume that the usual hypotheses hold.

Definition 2.1.2. A random variable T: $\Omega \to [0, \infty]$ is a stopping time if the event $\{T \leq t\} \in \mathcal{F}_t$, for every t, $0 \leq t < \infty$.

One important consequence of the right continuity of the filtration is the following theorem:

Theorem 2.1.1. *The event $\{T < t\} \in \mathcal{F}_t$, $0 \leq t \leq \infty$, if and only if T is a stopping time.*

A stochastic process X on $(\Omega, \mathcal{F}, P)$ is a collection of random variables $(X_t)_{0 \leq t \leq \infty}$. The process X is said to be adapted if $X_t \in \mathcal{F}_t$ (that is, $\mathcal{F}_t$-measurable) for each t.

Definition 2.1.3. A stochastic process X is said to be *cádlág* if it a.s. has sample paths which are right continuous, with left limits.

Let $(T_n)_{n \geq 0}$ be a strictly increasing sequence of positive random variables. We always take $T_0 = 0$ a.s. Recall that the indicator function $1_{\{t \geq T_n\}}$ is defined as:

$$1_{\{t \geq T_n\}}(\omega) = \begin{cases} 1 & \text{if } t \geq T_n(\omega) \\ 0 & \text{if } t < T_n(\omega). \end{cases}$$

Definition 2.1.4. The process $N = (N_t)_{0 \leq t \leq \infty}$ defined by

$$N_t = \sum_{n \geq 1} 1_{\{t \geq T_n\}},$$

with values in $\mathbb{N} \cup \{\infty\}$ ($\mathbb{N} = \{0, 1, 2, \cdots\}$) is called the counting process associated to the sequence $(T_n)_{n \geq 1}$.

Set $T = \sup_n T_n$, then the random variable T is the explosion time of N. If $T = \infty$ a.s., then N is a counting process without explosions. Note that for $0 \leq s < t < \infty$ we have

$$N_t - N_s = \sum_{n \geq 1} 1_{\{s < T_n \leq t\}}.$$

The increment $N_t - N_s$ counts the number of random times T_n that occur between the fixed times s and t.

Theorem 2.1.2. *A counting process N is adapted if and only if the associated random variables $(T_n)_{n \geq 1}$ are stopping times.*

Note that a counting process without explosions has right continuous paths with left limits; hence a counting process without explosions is cádlág.

Definition 2.1.5. An adapted counting process N without explosions is a Poisson process if
(i) for any s, t, $0 \leq s < t < \infty$, $N_t - N_s$ is independent of $\mathcal{F}_s$;
(ii) for any s, t, u, v, $0 \leq s < t < \infty$, $0 \leq u < v < \infty$, $t - s = v - u$, then the distribution of $N_t - N_s$ is the same as that of $N_v - N_u$.

Properties (i) and (ii) are known respectively as increments independent of the past, and stationary increments.

Definition 2.1.6. An adapted process $W = (W_t)_{0 \leq t \leq \infty}$ taking values in R^n is called an n-dimensional Brownian motion if:
(i) for $0 \leq s < t < \infty$, $W_t - W_s$ is independent of $\mathcal{F}_s$ (increments are independent of the past);
(ii) for $0 < s < t$, $W_t - W_s$ is a Gaussian random variable with mean zero and variance matrix $(t - s)C$, for a given, non random matrix C.

The Brownian motion starts at x if $P(B_0 = x) = 1$. The existence of Brownian motion is proved using a path-space construction, together with Kolmogorov's extension theorem.

Definition 2.1.7. A process $\{X_t,\ t \in T\}$ is said to be a Markov process if for any increasing collection $t_1, t_2, \cdots, t_n$ in T

$$P(X_{t_n} < x_n | X_{t_\nu} = x_\nu, \nu = 1, \cdots, n-1) = P(X_{t_n} | X_{t_{n-1}} = x_{n-1}).$$

The finite-dimensional distribution of a Markov process $\{X_t,\ t \in T\}$ can be completely determined by the initial and transitional distributions.

Definition 2.1.8. An adapted process $X = (X_t)_{0 \le t < \infty}$ with $X_0 = 0$ a.s. is a Lévy process if
(i) X has increments independent of the past: that is, $X_t - X_s$ is independent of $\mathcal{F}_s$, $0 \le s < t < \infty$;
(ii) X has stationary increments: that is, $X_t - X_s$ has the same distribution as X_{t-s}, $0 \le s < t < \infty$;
(iii) X_t is continuous in probability: that is, $\lim_{t \to s} X_t = X_s$, where the limit is taken in probability.

Definition 2.1.9. A real valued, adapted process $X = (X_t)_{0 \le t < \infty}$ is called a martingale with respect to the filtration $(\mathcal{F}_t)_{0 \le t < \infty}$ if
(i) $X_t \in L^1(dP)$; that is, $E\{\mid X_t \mid\} < \infty$;
(ii) if $s \le t$, then $E\{X_t \mid \mathcal{F}_s\} = X_s$, a.s.

Let $W = (W_t)_{0 \le t \le \infty}$ be a one dimensional standard Brownian motion with $W_0 = 0$. Then $M_t = W_t^2 - t$ is a martingale.

Let N be a Poisson process with intensity λ. Then $n_t \triangleq N_t - \lambda t$ and $(N_t - \lambda t)^2 - \lambda t$ are martingales.

Definition 2.1.10. An adapted process $X = \{X_t\}$, $t \geq 0$, is a semimartingale if it has a decomposition of the form

$$X_t = X_0 + M_t + A_t,$$

with $M_0 = A_0 = 0$, M a local martingale with jumps bounded by β, where $\beta > 0$, A is of integrable variation.

Definition 2.1.11. A is of integrable variation if $E\{\int_0^\infty \mid dA_s \mid\} < \infty$.

The quadratic variation process of a semimartingale, also known as the bracket process, is a simple object that nevertheless plays a fundamental role.

Definition 2.1.12. Let X, Y be semimartingales. The quadratic variation process of X, denoted $[X, X] = ([X, X]_t)_{t \geq 0}$, is defined by:

$$[X, X] = X^2 - 2 \int X_- dX,$$

(recall: $X_{0-} = 0$); The quadratic covariation of X, Y, also called the bracket process of X, Y, is defined by:

$$[X, Y] = XY - \int X_- dY - \int Y_- dX.$$

It is clear that the operation $(X, Y) \to [X, Y]$ is bilinear and symmetric. We therefore have a Polarization identity:

$$[X, Y] = \frac{1}{2}([X + Y, X + Y] - [X, X] - [Y, Y]).$$

Definition 2.1.13. For a semimartingale X, the process $[X, X]^c$ denotes the path by continuous part of $[X, X]$.

We can then write:

$$[X, X]_t = [X, X]_t^c + X_0^2 + \sum_{0 < s \leq t} (\Delta X_s)^2,$$

and

$$[X, X]_t = [X, X]_t^c + \sum_{0 \leq s \leq t} (\Delta X_s)^2,$$

where $\Delta X_s = X_s - X_{s-}$. Observe that $[X, X]_0^c = 0$.

Definition 2.1.14. A semimartingale X will be called quadratic pure jump if $[X,X]^c = 0$.

If X is quadratic pure jump, then $[X,X]_t = X_0 + \sum_{0<s\leq t}(\Delta X_s)^2$. The Poisson process N is an obvious example of a quadratic pure jump semimartingale. From the definition, we see that $[N,N]_t = N_t$. If ϕ is a progressively measurable process then we compute that

$$\int_0^t \phi_s dN_s = \sum_{\Delta N_s=1, s\leq t} \phi_s,$$

and

$$\int_0^t \phi_s dn_s = \sum_{\Delta N_s=1, s\leq t} \phi_s - \lambda \int_0^t \phi_s ds.$$

For example,

$$\int_0^t N_s dN_s = \sum_{\Delta N_s=1, s\leq t} N_s = 1+2+\cdots+N_t = \frac{N_t(N_t+1)}{2},$$

and, since ΔN_s=0 for all but countably many values of s,

$$\int_0^t \Delta N_s dn_s = \sum_{\Delta N_s=1, s\leq t} \Delta N_s - \int_0^t \Delta N_s ds = N_t.$$

Equation above shows that the integral of a bounded progressively measurable (in particular, adapted) random process with respect to a martingale is not necessarily a martingale. The reason is that the integrand and integrator can simultaneously jump in a positively correlated way. Integrals with respect to martingales are martingales if the integrands are "predictable", such as $\int_0^t N_{s-} dn_s$ is martingale.

Definition 2.1.15. Two probability laws P, Q on $(\Omega, \mathcal{F})$ are said to be equivalent if $P \ll Q$ and $Q \ll P$. (Recall that $P \ll Q$ denotes that P is absolutely continuous with respect to Q.) We write $Q \sim P$.

If $Q \ll P$, then there exits a random variable Z in $L^1(dP)$ such that $\frac{dQ}{dP} = Z$ and $E_P\{Z\} = 1$, where E_P denotes expectation with respect to the law P. We let

$$Z_t = E_P\{\frac{dQ}{dP}|\mathcal{F}_t\}$$

be the right continuous version. Then Z is a uniformly integrable cádlág martingale and hence a semimartingale. Note that if Q is equivalent to P, then $\frac{dP}{dQ} \in L^1(dQ)$ and $\frac{dP}{dQ} = (\frac{dQ}{dP})^{-1}$.

2.2 Fundamental Theory and Some Examples

Theorem 2.2.1. *(Generalized Itó Formula) Let f be a twice continuously differentiable function on R and let X be a real semimartigale. Then $f(X_t)$ is also a semimartingale and*

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_{s-})dX_s + \frac{1}{2}\int_0^t f''(X_{s-})d[X,X]_s^c \\ &\quad + \sum_{0<s\leq t}\{f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\}. \end{aligned} \tag{2.2.1}$$

Theorem 2.2.2. *(The Multi-dimensional Generalized Itó Formula) Suppose X is a process with values in R^n, each of whose components X^i is a semimartingale. Suppose F is a real valued twice continuously differentiable function on R^n. Then $F(X_t)$ is a semimartingale and, with equality denoting indistinguishability:*

$$\begin{aligned} F(X_t) &= F(X_0) + \sum_{i=1}^n \int_0^t \frac{\partial}{\partial x_i}F(X_{s-})dX_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2}{\partial x_i x_j}F(X_{s-})d[X^i, X^j]_s^c) \\ &\quad + \sum_{0<s\leq t}\{F(X_s) - F(X_{s-}) - \sum_{i=1}^n \frac{\partial}{\partial x_i}F(X_{s-})\Delta X_s^i\}. \end{aligned} \tag{2.2.2}$$

Example: We consider Merton's Model (Merton [59]):

$$dS_t = S_t\mu dt + S_t\sigma dW_t + S_{t-}d(\sum_{i=1}^{N(t)} Y_i),$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, W_t is a standard Brownian motion, $N(t)$ is a standard Poisson process independent of W_t, and $\{Y_i\}_{i=1}^{\infty}$ are random variables independent of $(N(t), W_t)$.

Letting:

$$f(S_t) = \log S_t.$$

Using generalized Itó Formula (2.2.1), we have:

$$\begin{aligned} \log S_t &= \log S_0 + \int_0^t \frac{1}{S_{r^-}} dS_r + \frac{1}{2}\int_0^t (-\frac{1}{S_{r^-}^2}) d[S,S]_r^c \\ &\quad + \sum_{0<r\leq t} \{\log S_r - \log S_{r^-} - \frac{1}{S_{r^-}} \Delta S_r\}. \end{aligned} \tag{2.2.3}$$

Note that:

$$S_r = S_{r^-}(1 + Y_{N(r)}),$$

from (2.2.3), we have:

$$\begin{aligned} \log S_t &= \log S_0 + \int_0^t \{(\mu - \frac{1}{2}\sigma^2)dr + \sigma dW_r + d(\sum_{i=1}^{N(r)} Y_i)\} \\ &\quad + \sum_{i=1}^{N(t)} \{\log(1+Y_i) - Y_i\}, \end{aligned} \tag{2.2.4}$$

namely,

$$\log S_t = \log S_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t + \sum_{i=1}^{N(t)} \log(1+Y_i), \tag{2.2.5}$$

so

$$\begin{aligned} S_t &= S_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t + \sum_{i=1}^{N(t)} \log(1+Y_i)) \\ &= S_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t) \Pi_{i=1}^{N(t)} (1+Y_i). \end{aligned} \tag{2.2.6}$$

We give another example: Let N_t^1 and N_t^2 be Poisson processes with rates λ^1 and λ^2, respectively. We suppose the exchange rate is now given by the equation:

$$\frac{dS_t}{S_{t-}} = \mu_t dt + \sigma_t dW_t + \alpha dN_t^1 - \beta dN_t^2.$$

Here W_t is a standard Brownian motion, μ_t is drift and σ_t is volatility. We also assume N_t^1 and N_t^2 do not have common jump times. We want to calculate $d(\frac{1}{S_t})$.

Using generalized Itó Formula (2.2.1), we have:

$$d(\frac{1}{S_t}) = -\frac{1}{S_{t-}^2}dS_t + \frac{1}{S_{t-}}\sigma_t^2 dt + \sum_{0<r\leq t}\{\frac{1}{S_r} - \frac{1}{S_{r-}} + \frac{1}{S_{r-}}(\alpha\Delta N_r^1 - \beta\Delta N_r^2)\},$$

i.e

$$d(\frac{1}{S_t}) = -\frac{1}{S_{t-}}(\mu_t - \sigma_t^2)dt - \frac{1}{S_{t-}}\sigma_t dW_t + \frac{\alpha^2}{1+\alpha}\frac{1}{S_{t-}}dN_t^1 + \frac{\beta^2}{1-\beta}\frac{1}{S_{t-}}dN_t^2. \quad (2.2.7)$$

Equation (2.2.7) is the corrected version of the equation on page 55, line 5 in Chesney and Elliott [14].

Corollary 2.2.3. *If X and Y are semimartingales then the product XY is a semimartingale and*

$$X_t Y_t = \int_0^t X_{s-} dY_s + \int_0^t Y_{s-} dX_s + [X,Y]_t. \quad (2.2.8)$$

That is, in differential form:

$$d(XY)_t = X_{t-} dY_t + Y_{t-} dX_t + d[X,Y]_t. \quad (2.2.9)$$

Theorem 2.2.4. *(Doléans-Dade Exponential Formula) Let X_t be a semimartingale with $X_0 = 0$. Then there exists a (unique) semimartingale Z that satisfies the equation $Z_t = Z_0 + \int_0^t Z_{s-} dX_s$, and Z is given by*

$$Z_t = Z_0 \exp(X_t - \frac{1}{2}[X,X]_t^c)\Pi_{0<s\leq t}(1+\Delta X_s)\exp(-\Delta X_s),$$

where the infinite product converges.

Example: We consider Merton's Model (Merton [59]) again:

$$\begin{aligned} dS_t &= S_t\mu dt + S_t\sigma dW_t + S_{t-}d(\sum_{i=1}^{N(t)} Y_i) \\ &= S_{t-}(\mu dt + \sigma dW_t + d(\sum_{i=1}^{N(t)} Y_i)). \end{aligned}$$

Letting:

$$dS_t = S_{t-}dX_t.$$

We find that:

$$dX_t = \mu dt + \sigma dW_t + d(\sum_{i=1}^{N(t)} Y_i)).$$

Hence

$$X_t = \mu t + \sigma W_t + \sum_{i=1}^{N(t)} Y_i,$$

and

$$\begin{aligned} S_t &= S_0 \exp(X_t - \frac{1}{2}[X,X]^c{}_t)\Pi_{0<r\leq t}(1+\Delta X_r)\exp(-\Delta X_r) \\ &= S_0 \exp(\mu t + \sigma W_t + \sum_{i=1}^{N(t)} Y_i - \frac{1}{2}\sigma^2 t)\Pi_{i=1}^{N(t)}(1+Y_i)\exp(-Y_i) \\ &= S_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t)\Pi_{i=1}^{N(t)}(1+Y_i). \end{aligned} \tag{2.2.10}$$

Definition 2.2.1. For a semimartingale X, with $X_0 = 0$ the stochastic exponential of X, written $\mathcal{E}(X)$, is the (unique) semimartingale Z that is a solution of: $Z_t = 1 + \int_0^t Z_{s-}dX_s$.

Corollary 2.2.5. *If X and Y are semimartingales with $X_0 = Y_0 = 0$. Then*

$$\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + [X,Y]).$$

Theorem 2.2.6. *(Lévy-Khintchine formula) Let X be a Lévy process with Lévy measure ν. Then*

$$E\{e^{iuX_t}\} = e^{-t\psi(u)},$$

where

$$\psi(u) = \frac{\sigma^2}{2}u^2 - i\alpha u + \int_{\{|x|\geq 1\}}(1 - e^{iux})\nu(dx) + \int_{\{|x|<1\}}(1 - e^{iux} + iux)\nu(dx).$$

Moreover given ν,σ^2,α, the corresponding Lévy process is unique in distribution.

Girsanov Transform theorem is fundamental in the general theory of stochastic analysis. It is also important in many applications, for examples in finance and filtering. In fact, the law of the new process will be absolutely continuous with respect to the law of the original process and we can compute explicitly the Radon-Nikodym derivative.

Theorem 2.2.7. *Let $Y(t) \in R^n$ be an Itó process of the form*

$$dY(t) = a(t,w)dt + dB(t) \qquad t \leq T, \quad Y_0 = 0.$$

Where $T \leq \infty$ is a given constant and $B(t)$ is n-dimensional Brownian motion w.r.t P. Put

$$M_t = \exp(-\int_0^t a(s,w)dW_t - \frac{1}{2}\int_0^t a^2(s,w)ds); \quad t \leq T.$$

Assume that $a(s,w)$ satisfies Novikov's condition

$$E[\exp(\frac{1}{2}\int_0^T a^2(s,w)ds)] < \infty,$$

where $E = E_P$ is the expectation w.r.t P. Define the measure Q on $(\Omega, \mathcal{F}_T^{(n)})$ by

$$dQ(w) = M_T(w)dP(w).$$

Then $Y(t)$ is an n-dimensional Brownian motion w.r.t. the probability law Q, for $t \leq T$.

Remark: Theorem states that for all Borel sets $F_1, \cdots, F_k \in R^n$ and all $t_1, t_2, \cdots, t_k \leq T, k = 1, 2, \cdots$ we have:

$$Q[Y(t_1) \in F_1, \cdots, Y(t_k) \in F_k] = P[B(t_1) \in F_1, \cdots, B(t_k) \in F_k].$$

It is to say that $Q \ll P$ (Q is absolutely continuous w.r.t. P) with Radon-Nikodym derivative,

$$\frac{dQ}{dP} = M_T \qquad \text{on } \mathcal{F}_T^{(n)}.$$

Note that $M_T(w) > 0$ a.s., so we also have that $P \ll Q$. Hence the two measures Q and P are equivalent. Therefore we get

$$\begin{aligned} & P[Y(t_1) \in F_1, \cdots, Y(t_k) \in F_k] > 0 \\ \iff & Q[Y(t_1) \in F_1, \cdots, Y(t_k) \in F_k] > 0 \\ \iff & Q[B(t_1) \in F_1, \cdots, B(t_k) \in F_k] > 0; \quad t_1, \cdots, t_k \in [0, T]. \end{aligned}$$

Theorem 2.2.8. *Change of Measure (Conditional Bayes Theorem) Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-field. Suppose $\bar{P}$ is another probability measure, absolutely continuous with respect to P and with Radon-Nikodym derivative:*

$$\frac{d\bar{P}}{dP} \mid_{\mathcal{G}} = \Lambda.$$

If ψ is any $\bar{P}$ integrable random variable, then

$$\bar{E}[\psi|\mathcal{G}] = \frac{E[\Lambda\psi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]},$$

if $E[\Lambda|\mathcal{G}] > 0$ otherwise 0.

We shall adopt the following notation and definition:

If $\mathcal{M}$ is some family of processes, then $\mathcal{M}_{\text{loc}}$ will denote the family of processes

which are locally in $\mathcal{M}$. That is, $\{Y_t\} \in \mathcal{M}_{\text{loc}}$ if there is an increasing sequence of stopping times $\{T_n\}$ such that $\lim_n T_n = \infty$ a.s. and such that each stopped process $\{Y_t^{T_n}\} = \{Y_{t \wedge T_n}\}$ is in $\mathcal{M}$

Theorem 2.2.9. *Let N_t, $t \geq 0$, be a standard Poisson process with respect to $(P, \mathcal{F}_t)$ and let $\mathcal{F}_t = \sigma(N_s, s \leq t)$. Then there is a probability measure Q equivalent to P such that N_t has a unique decomposition*

$$N_t = Z_t + \int_0^t \lambda_s ds,$$

with respect to $(Q, \mathcal{F}_t)$, where $Z \in \mu_{loc}(Q, \mathcal{F}_t)$, λ is a positive predictable process, and $[Z, Z]_t = \int_0^t \lambda_s ds$. Furthermore, the likelihood ratio is given by

$$\Lambda_t = E(\frac{dQ}{dP}|\mathcal{F}_t) = (\Pi_{s \leq t} \lambda_s) \exp(- \int_0^t (\lambda_s - 1)).$$

In the following example, we give the estimation of stochastic volatility. In the Black-Scholes model, the stock volatility is usually constant. Many authors have proposed that the option prices are correlated with stochastic volatility. Scott [71], Hull and White [44], and Wiggins [75] generalized the model to allow stochastic volatility, and Melino and Turnbull [58] reported that this approach was successful in explaining the prices of currency options.

This example is a direct extension of Elliott, Hoek and Valencia [33]. They assumed that volatility is uncorrelated with spot return, but it can not explain the important skewness effects that arise from such correction in Hull and White [44]. In our work, we assume that volatility is corrected with spot return.

Suppose a price process S evolves in discrete time, $t = 0, 1, 2, \cdots$, with dynamics:

$$S_t = S_{t-1} \exp(\mu - \frac{1}{2}{\sigma_t}^2 + \sigma_t w_t^1),$$

here $\mu \in \mathbb{R}$ is the drift, w_t^1 is a sequence of i.i.d $N(0,1)$ random variables, and σ_t represents the volatility of the price change between time $(t-1)$ and t. The price sequence $S_0, S_1, \cdots$ is observed as are the logarithmic increments

$$y_t = \log \frac{S_t}{S_{t-1}} = \mu - \frac{1}{2}{\sigma_t}^2 + \sigma_t w_t^1.$$

We suppose that $\log \sigma_t$ has dynamics: $\log \sigma_t = a + b\log \sigma_{t-1} + \theta w_t^2$, here again, $\{w_t^2\}$ is a sequence of i.i.d $N(0,1)$ random variables and $a, b, \theta \in \mathbb{R}$. Writing $x_t = \log \sigma_t$, so that $\sigma_t = e^{x_t}$, we see:

$$\begin{aligned} x_t &= a + bx_{t-1} + \theta w_t^2 \\ y_t &= \mu - \frac{1}{2}e^{2x_t} + e^{x_t} w_t^1. \end{aligned}$$

Next, one lets

$$\begin{aligned} E(w_t^1 w_t^2) &= \rho \\ p(w_t^1, w_t^2) &= \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\{-\frac{1}{2(1-\rho^2)}((w_t^1)^2 + (w_t^2)^2 - 2\rho w_t^1 w_t^2)\}, \end{aligned}$$

where $\rho \neq 1$ and -1. Using variable transformation, we let:

$$\begin{aligned} w_t^1 &= (1-\rho^2)^{\frac{1}{2}} v_t + \rho u_t, \\ w_t^2 &= u_t. \end{aligned}$$

We have: $Eu_t = 0$, $Ev_t = 0$, and $E(u_t v_t) = 0$, then the models are:

$$x_t = a + bx_{t-1} + \theta u_t \tag{2.2.11}$$

$$y_t = \mu - \frac{1}{2}e^{2x_t} + (1-\rho^2)^{\frac{1}{2}} e^{x_t} v_t + \rho e^{x_t} u_t. \tag{2.2.12}$$

Reference Probability Methods:

The reference probability methods refer to a procedure where a probability measure change is introduced to reformulate the original process in a fictitious world so that well-known results can be applied. Then, the results are reinterpreted back to the real world with the original probability measure. The idea of using reference probability method in filtering goes back to classical work of Kalman and Bucy.

Let us assume that under the reference probability $\bar{P}$ both $\{x_t\}, \{y_t\}$ are sequences of i.i.d $N(0,1)$ random variable, write $\phi(\cdot)$ for the $N(0,1)$ density.

We define the σ-fields:

$$\mathcal{G}_t = \sigma\{x_0, x_1, x_2, \cdots, x_t, y_0, y_1, y_2, \cdots, y_t\};$$
$$\mathcal{Y}_t = \sigma\{y_0, y_1, y_2, \cdots, y_t\}.$$

Thus $\mathcal{G}_t$ is the complete filtration generated by the x and y sequences and $\mathcal{Y}_t$ is the complete filtration generated by the observations y.

Let

$$\lambda_k = \frac{\theta\phi(x_k)}{\phi(u_k)} \frac{e^{x_k}(1-\rho^2)^{\frac{1}{2}}\phi(y_k)}{\phi(v_k)},$$

for $k = 1, 2, \cdots$ where

$$\begin{aligned} u_k &= \theta^{-1}(x_k - a - bx_{k-1}) \\ v_k &= (1-\rho^2)^{\frac{1}{2}} e^{-x_k}(y_k - \mu + \frac{1}{2}e^{2x_k} - \rho e^{x_k}\theta^{-1}(x_k - a - bx_{k-1})). \end{aligned}$$

Set

$$\begin{aligned} \lambda_0 &= \frac{e^{x_0}\phi(y_0)(1-\rho^2)^{\frac{1}{2}}}{\phi(e^{-x_0}(1-\rho^2)^{\frac{1}{2}}(y_0 - \mu + \frac{1}{2}e^{2x_0}))} \\ \Lambda_t &= \prod_{k=0}^{t} \lambda_k. \end{aligned}$$

Define a new probability $\bar{P}$, by setting

$$\frac{d\bar{P}}{dP} \mid_{\mathcal{G}_t} = \Lambda_t.$$

We shall work under $\bar{P}$. The existence of $\bar{P}$ follows from Kolmogorov's extension theorem (Elliott [30]).

Lemma 2.2.10. *Under $\bar{P}$ the $\{x_t\}, \{y_t\}, t = 0, 1, 2, \cdots$ are sequences of i.i.d $N(0,1)$ random variables. Note x_t and y_t are independent of each other as well.*

Proof:

$$\begin{aligned}
\bar{P}(y_t \leq l | \mathcal{G}_{t-1}) &= \bar{E}[I(y_t \leq l) | \mathcal{G}_{t-1}] \\
&= \frac{E[\Lambda_t I(y_t \leq l) | \mathcal{G}_{t-1}]}{E[\Lambda_t | \mathcal{G}_{t-1}]} \\
&= \frac{E[\lambda_t I(y_t \leq l) | \mathcal{G}_{t-1}]}{E[\lambda_t | \mathcal{G}_{t-1}]}.
\end{aligned}$$

$$\begin{aligned}
&E[\lambda_t | \mathcal{G}_{t-1}] = E\{\frac{\theta\phi(x_t)}{\phi(u_t)} \frac{e^{x_t}(1-\rho^2)^{\frac{1}{2}}\phi(y_t)}{\phi(v_t)} | \mathcal{G}_{t-1}\} \\
&= E\{E\{\frac{\theta\phi(x_t)}{\phi(u_t)} \frac{e^{x_t}(1-\rho^2)^{\frac{1}{2}}\phi(y_t)}{\phi(v_t)} | \mathcal{G}_{t-1}, u_t\} | \mathcal{G}_{t-1}\} \\
&= E\{\frac{\theta\phi(a + bx_{t-1} + \theta u_t)}{\phi(u_t)} \\
&\quad \times E\{\frac{e^{x_t}(1-\rho^2)^{\frac{1}{2}}\phi(\mu - \frac{1}{2}e^{2x_t} + e^{x_t}(1-\rho^2)^{\frac{1}{2}}v_t + \rho e^{x_t}u_t)}{\phi(v_t)} | \mathcal{G}_{t-1}, u_t\} | \mathcal{G}_{t-1}\} \\
&= E\{\frac{\theta\phi(a + bx_{t-1} + \theta u_t)}{\phi(u_t)} \\
&\quad \times \int_{-\infty}^{\infty} \frac{e^{x_t}(1-\rho^2)^{\frac{1}{2}}\phi(\mu - \frac{1}{2}e^{2x_t} + e^{x_t}(1-\rho^2)^{\frac{1}{2}}v_t + \rho e^{x_t}u_t)}{\phi(v_t)} \times \phi(v_t)dv_t | \mathcal{G}_{t-1}\} \\
&= \int_{-\infty}^{\infty} \phi(y_t)dy_t \times E\{\frac{\theta\phi(a + bx_{t-1} + \theta u_t)}{\phi(u_t)} | \mathcal{G}_{t-1}\} = 1.
\end{aligned}$$

$$E[\lambda_t I(y_t \le l)|\mathcal{G}_{t-1}] = E\{\frac{\theta\phi(x_t)}{\phi(u_t)}\frac{e^{x_t}(1-\rho^2)^{\frac{1}{2}}\phi(y_t)}{\phi(v_t)}I(y_t \le l)|\mathcal{G}_{t-1}\}$$

$$= E\{\frac{\theta\phi(a+bx_{t-1}+\theta u_t)}{\phi(u_t)}E\{\frac{e^{x_t}(1-\rho^2)^{\frac{1}{2}}\phi(\mu-\frac{1}{2}e^{2x_t}+e^{x_t}(1-\rho^2)^{\frac{1}{2}}v_t+\rho e^{x_t}u_t)}{\phi(v_t)}$$

$$\times I(y_t \le l)|\mathcal{G}_{t-1}, u_t\}|\mathcal{G}_{t-1}\}$$

$$= E\{\frac{\theta\phi(a+bx_{t-1}+\theta u_t)}{\phi(u_t)} \times \int_{-\infty}^{\infty}\frac{e^{x_t}(1-\rho^2)\phi(\mu-\frac{1}{2}e^{2x_t}+e^{x_t}(1-\rho^2)^{\frac{1}{2}}v_t+\rho e^{x_t}u_t)}{\phi(v_t)}$$

$$\times I(\mu-\frac{1}{2}e^{2x_t}+e^{x_t}(1-\rho^2)^{\frac{1}{2}}v_t+\rho e^{x_t}u_t \le l)\phi(v_t)dv_t|\mathcal{G}_{t-1}\}$$

$$= \int_{-\infty}^{\infty} I(y_t \le l)\phi(y_t)dy_t \times E\{\frac{\theta\phi(a+bx_{t-1}+\theta u_t)}{\phi(u_t)}|\mathcal{G}_{t-1}\}$$

$$= \int_{-\infty}^{l}\phi(y_t)dy_t = \bar{P}(y_t \le l).$$

$$\bar{P}(y_t \le l|\mathcal{G}_{t-1}) = \bar{P}(y_t \le l).$$

The same as $\bar{P}(x_t \le l|\mathcal{G}_{t-1}) = \bar{P}(x_t \le l)$.

Remark: Conversely, We suppose we start with a probability measure $\bar{P}$ on $(\Omega, \mathcal{F})$ such that under $\bar{P}$ both $\{x_t\}$, $\{y_t\}$ are sequences of i.i.d. $N(0,1)$ random variable. we then wish to construct a probability measure P, such that under P both $\{v_t\}$, $\{u_t\}$ are a sequence of i.i.d. $N(0,1)$ random variables. To contract P from $\bar{P}$, we introduce the inverse $\bar{\lambda}_t = \lambda_t^{-1}$ and $\bar{\Lambda}_t = \Lambda_t^{-1}$, $\bar{\Lambda}_0 = 1$ and define P by setting $\frac{dP}{d\bar{P}}|_{\mathcal{G}_t} = \bar{\Lambda}_t$.

Lemma 2.2.11. *Under P the $\{v_t\}, \{u_t\}, t = 0, 1, 2, \cdots$ are sequences of i.i.d $N(0,1)$ random variables.*

Proof: The proof is the similar to that of lemma 2.2.10. also *see* [35].

We shall again use Bayes' theorem, for any *Borel* measurable function f,

$$E[f(x_t)|\mathcal{Y}_t] = \frac{\bar{E}[\bar{\Lambda}_t f(x_t)|\mathcal{Y}_t]}{\bar{E}[\bar{\Lambda}_t|\mathcal{Y}_t]}.$$

The numerator defines a measure; suppose it has a density $q(\cdot)$ so that

$$\bar{E}[\bar{\Lambda}_t I(x_t \in dz)|\mathcal{Y}_t] = q_t(z)dz.$$

Then

$$\bar{E}[\bar{\Lambda}_t f(x_t)|\mathcal{Y}_t] = \int_{-\infty}^{\infty} f(z)q_t(z)dz. \tag{2.2.13}$$

We now obtain a recursive update for $q_t(z)$.

$$\begin{aligned} q_t(z) &= \frac{1}{\theta e^z(1-\rho^2)^{\frac{1}{2}}\phi(y_t)} \int_{-\infty}^{\infty} \phi(e^{-z}(1-\rho^2)^{\frac{1}{2}}(y_t - \mu + \frac{1}{2}e^{2z} - \rho e^z \theta^{-1}(z-a-bx)) \\ &\quad \times \phi(\theta^{-1}(z-a-bx))q_{t-1}(x)dx. \end{aligned}$$

This gives the formula for updating the unnormalized conditional density for $x_t = \log \sigma_t$ given $\mathcal{Y}_t$.

Putting $f(x) = 1$, in (2.2.13) we have $\bar{E}[\bar{\Lambda}_t|\mathcal{Y}_t] = \int_{-\infty}^{\infty} q_t(z)dz$, so the normalized conditional density of $x_t = \log \sigma_t$ given $\mathcal{Y}_t$ is

$$p_t(z) = \frac{q_t(z)}{\int_{-\infty}^{\infty} q_t(x)dx}.$$

Furthermore, taking $f(x_t) = x_t$, we see

$$E[x_t|\mathcal{Y}_t] = \frac{\int_{-\infty}^{\infty} z q_t(z)dz}{\int_{-\infty}^{\infty} q_t(z)dz}.$$

This is the optimal estimate of the logarithm of the volatility given the observations of the prices.

Chapter 3

Numerical Methods and Monte Carlo Simulation

In this Chapter, we first introduce some simulation methods of, for example, the exponential random variable, and the normal distribution random variable. Secondly, we consider the numerical methods for diffusion with jump model and extend Milstein's result. Some sample paths of typical stochastic processes are also generated such as continuous time finite state Markov chain. All of them are the basis for the simulation study Chapter 5.

3.1 Basic Sampling Methods and Generating Sample Paths

In this section we shall assume that we have a subroutine RANDOM which provides us with $U(0,1)$ uniformly distributed on $[0,1]$ pseudo-random numbers. We shall see how we can then use this subroutine to generate pseudo-random numbers with other commonly encountered distributions, in particular those simulated in Chapter 5.

A two-point random variable X taking values $x_1 < x_2$ with probabilities p_1 and

$p_2 = 1 - p_1$ can be generated easily from a uniformly distribution on $[0,1]$ a $U(0,1)$ random variable U with

$$X = \begin{cases} x_1, & 0 \le U \le p_1 \\ x_2, & p_1 < U \le 1 \end{cases}$$

This idea extends readily to an M-state random variable X taking values $x_1 < x_2 < \cdots < x_M$ with nonzero probabilities $p_1, p_2, \cdots, p_M$ where $\sum_{i=1}^{M} p_i = 1$. With $s_0 = 0$ and $s_j = \sum_{i=1}^{j} p_i$ for $j = 1, 2, \cdots, M$ we set $X = x_{j+1}$ if $s_j < U \le s_{j+1}$ for $j = 0, 1, 2, \cdots, M-1$.

For a continuous random variable X the corresponding method requires the probability distribution function F_X to be inverted when this is possible. For a number $0 < U < 1$ we define $x(U)$ by $U = F_X(x(U))$, so $x(U) = F_X^{-1}(U)$ if F_X^{-1} exists, or in general

$$x(U) = \inf\{x : F_X(x) \ge U\}. \tag{3.1.1}$$

This is called the inverse transform method and is best used when (3.1.1) is easy to evaluate. For example: Exponential random variable with parameter $\theta > 0$ and distribution $F_X(x)$ given:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp(-\theta x), & x \ge 0 \end{cases}$$

has an invertible distribution function with

$$x(U) = F_X^{-1}(U) = -\log(1-U)/\theta \quad \text{for} \quad 0 < U < 1,$$

simplifies to

$$X = -\log(U)/\theta, \quad \text{for} \quad 0 < U < 1.$$

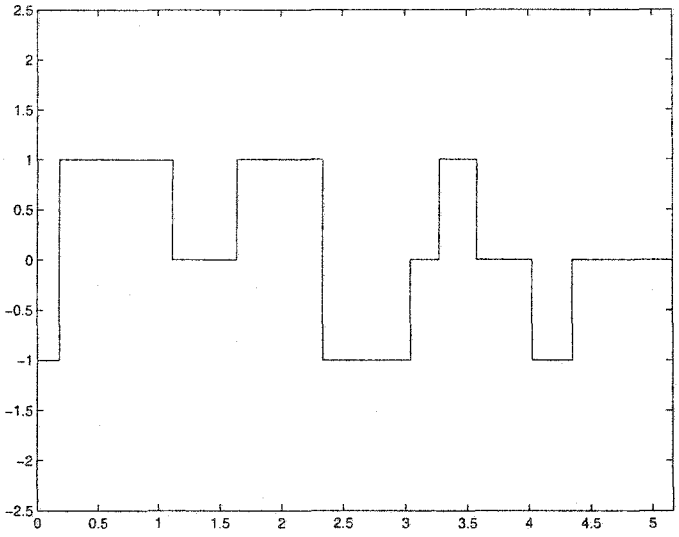

Figure 3.1: Three state continuous time Markov chain

Finite State Continuous Time Markov Chain: When we study the generators for Bernoulli, Binomial random variables and exponential random variables, we can handle finite state continuous time Markov chain. Let $X(t)$ be distributed over a finite state space $\chi = \{x_1, x_2, \cdots, x_M\}$ according to an M-dimensional probability vector $p(t)$ for each $t \geq 0$. Here we only consider the case when the Markov chain is homogenous, namely when the transition matrices $P(t_0; t_1)$ depend only on the time difference $t_1 - t_0$, that is $P(t_0; t_1) = P(0; t_1 - t_0)$ for all $0 \leq t_0 \leq t_1$, and write $P(t)$ for $P(0; t)$. There exists an $M \times M$ intensity matrix $A = (a^{i,j})$ with components

$$a^{i,j} = \begin{cases} \lim_{t \to 0} \frac{p^{i,j}(t)}{t}, & i \neq j \\ \lim_{t \to 0} \frac{p^{i,i}(t)-1}{t}, & i = j \end{cases}$$

which, together with the initial probability vector $p(0)$, characterizes completely the homogenous time Markov chain. Moreover, the waiting time of a homogenous continuous time Markov chain, that is the time between transitions from a state x_i to any other state, is exponentially distributed with intensity parameter $\lambda_i = \Sigma_{j \neq i} a^{i,j}$.

Figure 3.1 is the realization from a 3 state continuous time Markov chain. X takes values +1, 0, and -1 with the initial probability vector $p(0) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, for $t \geq 0$ has intensity matrix

$$A = \begin{bmatrix} -0.6 & 0.4 & 0.2 \\ 0.4 & -1 & 0.6 \\ 0.5 & 0.3 & -0.8 \end{bmatrix}.$$

When we use C/C++ package to write program we must write the normal distribution random number generator function by ourselves.

If $Z \sim N(0,1)$ and $X = \mu + \sigma Z$ then $X \sim N(\mu, \sigma^2)$; hence we just need a method for generating standard normal variables. we give two popular methods below:

Box-Muller method: If $Z \sim N(0, I)$ in $\mathbb{R}^2$ then

(i) $R = Z_1^2 + Z_2^2$ is exponentially distributed with mean 2, i.e.,

$$P(R \leq x) = 1 - e^{-x/2}.$$

(ii) Given R, then the point (Z_1, Z_2) is uniformly distributed on the circle of radius $\sqrt{R}$ centered at the origin.

Generate two independent Uniform variates U_1, $U_2 \sim$ U$[0,1]$.

$$R \leftarrow -2\log(U_1)$$

$$V \leftarrow 2\pi U_2$$

$$Z_1 \leftarrow \sqrt{R}\cos(V)$$

$$Z_2 \leftarrow \sqrt{R}\sin(V),$$

return Z_1, Z_2.

Inverse Cumulative Normal Distribution: Draw $U \sim U[0,1]$ and compute:

$$x = N^{-1}(U) \qquad \text{where} \quad N(z) = \frac{1}{2\pi}\int_{-\infty}^{z} e^{-v^2/2}dv,$$

then x is a Gaussian sample. To compute N^{-1}, Beasley and Springer [6] uses a rational approximation and take U uniform sample

$$\begin{aligned} v &= U - 0.5 \\ y &= v * v \\ N^{-1}(u) &= \frac{v * (a_3 y^3 + a_2 y^2 + a_1 y + a_0)}{b_3 y^4 + b_2 y^3 + b_1 y^2 + b_0 y + 1}. \end{aligned}$$

with

$a_0 = 2.50662823884$	$b_0 = -8.47351093090$
$a_1 = -18.61500062529$	$b_1 = 23.08336743743$
$a_2 = 41.3911977354$	$b_2 = -21.0622410101826$
$a_3 = -25.44106049637$	$b_3 = 3.13082909833$

Multivariate Normal Distribution: In many financial applications one has to generate variates according to a multivariate normal distribution with expected value μ and covariance matrix Σ. This task may be accomplished by obtaining the Cholesky factor for Σ, i.e., an upper triangular matrix L such that $\Sigma = L^T L$. Then we may apply the following algorithm:

(i) Generate n independent standard normal variaes $Z_1, \cdots, Z_n \sim N(0,1)$.

(ii) Return $X = \mu + L^T Z$, where $Z = [Z_1, \cdots, Z_n]^T$.

Wiener Processes: The Wiener processes were proposed by Norbert Wiener as a mathematical description of Brownian motion, the erratic motion of a grain of pollen

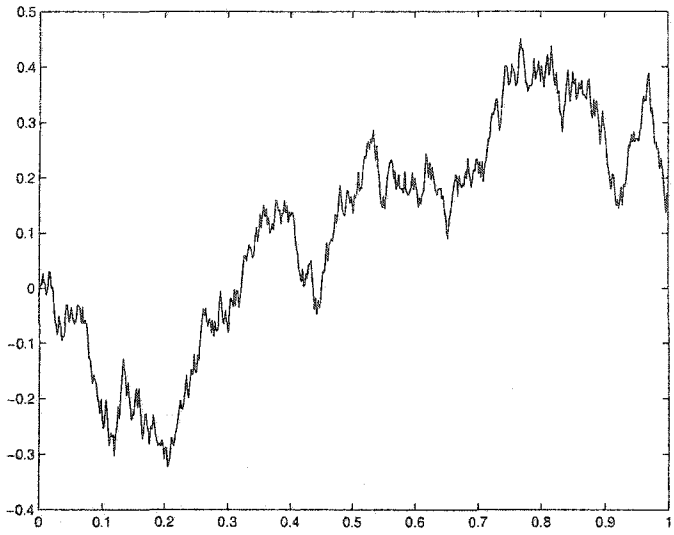

Figure 3.2: A linearly interpolated Wiener sample path

on a water surface due to its being continually bombarded by water molecules.

We define a standard Wiener process $W = \{W(t), t \geq 0\}$ to be a continuous Gaussian process with independent increments such that

$$W(0) = 0, \quad \text{with probability 1}, \quad E(W(t)) = 0, \quad \text{Var}(W(t) - W(s)) = t - s,$$

for all $0 \leq s \leq t$. According to this definition, $W(t) - W(s)$ is $N(0; t - s)$ Gaussian distributed for $0 \leq s < t$ and the increments $W(t_2) - W(t_1)$ and $W(t_4) - W(t_3)$ are independent for all $0 \leq t_1 < t_2 \leq t_3 < t_4$.

We generate and plot the linearly interpolated trajectory of a Wiener process on $[0, 1]$ at the time instants $t_k = k2^{-9}$ for $k = 0, 1, \cdots, 2^9$ using independent Gaussian increments $W(t_{k+1}) - W(t_k) \sim N(0, 2^{-9})$, see Figure 3.2.

Random Walks: We can approximate a standard Wiener process in distribution on any finite time interval by means of a scaled random walk. For example , we can

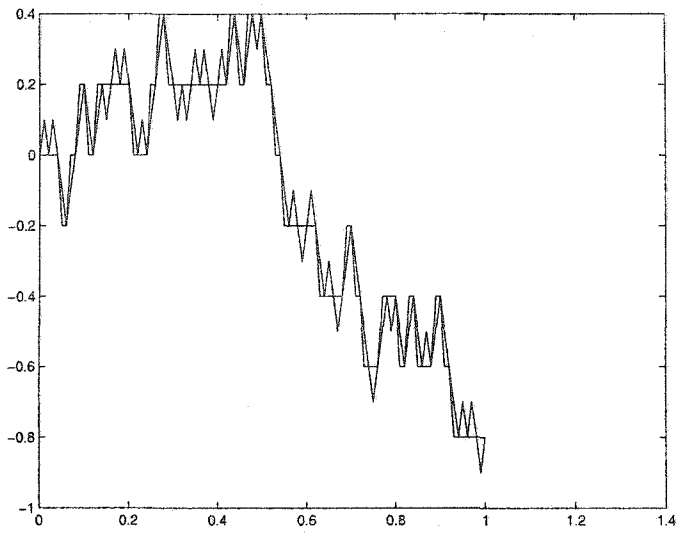

Figure 3.3: Sample paths of the random walk $S_{50}(t)$ and $S_{100}(t)$.

subdivide the unit interval $[0,1]$ into M subintervals

$$0 = t_0^{(M)} < t_1^{(M)} < \cdots < t_k^{(M)} < \cdots < t_M^{(M)} = 1,$$

of equal length $\Delta t = 1/M$ and construct a stepwise continuous random walk $S_M(t)$ by taking independent, equally probable steps of length $\pm\sqrt{\Delta t}$ at the end of each subinterval. We start with independent two-point random variables $\{X_i\}_{i=1}^k$ taking values ± 1 with equal probability and define

$$S_M(t_k^{(M)}) = (X_1 + X_2 + \cdots + X_k)\sqrt{\Delta t},$$

where we interpolate linearly by

$$S_M(t) = S_M(t_k^{(M)}) + \frac{t - t_k^{(M)}}{t_{k+1}^{(M)} - t_k^{(M)}}\left(S_M(t_{k+1}^{(M)}) - S_M(t_k^{(M)})\right)$$

on $t_k^{(M)} \le t < t_{k+1}^{(M)}$ for $k = 0, 1, \cdots, M-1$, where $S_M(0) = 0$.

Form a linearly interpolated random walk $S_{100}(t)$ on $[0,1]$ using the two-point random number generator and plot $S_{100}(t)$ against t. Repeat this for other sequences

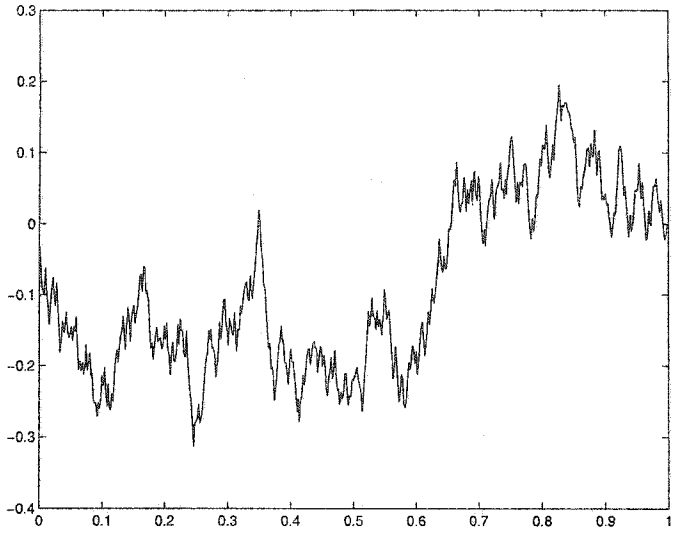

Figure 3.4: A path of a Brownian bridge process.

corresponding to different initial seeds and compare the plotted paths, see Figure 3.3.

Brownian Bridge Process: A useful modification of Wiener process has sample paths that all pass through the same initial point x, not necessarily 0, and a given point y at a later time $t = T$. This process $B_{0,x}^{T,y}$ is defined sample pathwise for $0 \leq t \leq T$ by

$$B_{0,x}^{T,y}(t,\omega) = x + W(t,\omega) - \frac{t}{T}\{W(T,\omega) - y + x\}$$

and is called a Brownian bridge or a tied-down Wiener process. It is a Gaussian process satisfying the constraints $B_{0,x}^{T,y}(0,\omega) = x$ and $B_{0,x}^{T,y}(T,\omega) = y$, so can be considered as a kind of conditional Wiener process. Since it is Gaussian it is determined uniquely by its means and covariances, which are

$$\mu(t) = x - \frac{t}{T}(x-y) \quad \text{and} \quad C(s,t) = \min\{s,t\} - \frac{st}{T},$$

for $0 \leq s,t \leq T$, respectively, see Figure 3.4.

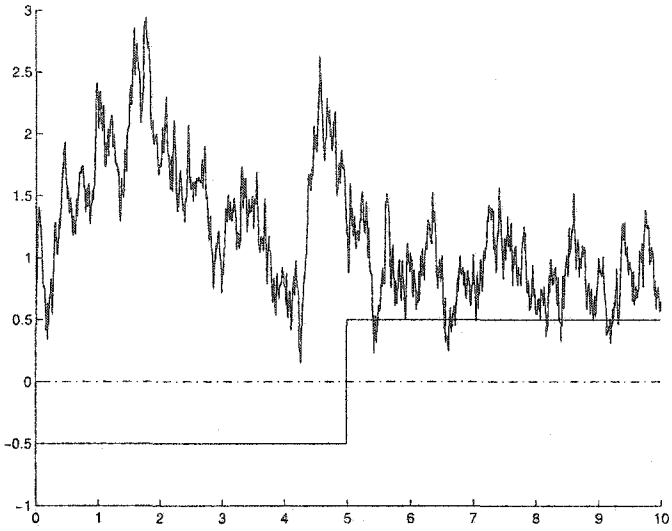

Figure 3.5: Paths of the Markov chain and Bessel process.

Bessel processes are an important family of diffusion processes which have applications in finance and other areas. Let us call a process $R = \{R_t, 0 \leq t\}$ a generalized Bessel process if it satisfies a stochastic differential equation of the type

$$dR_t = [\frac{\alpha_t}{R_t} + b_t R_t]dt + c_t dW_t,$$

for $t \geq 0$ with given initial value $R_0 \geq 0$. Here α, b and c represents externally given functions and $W = \{W_t, 0 \leq t\}$ is a standard Brownian motion on a given probability space$(\Omega, \mathcal{F}, P)$. From Itó's formula, we know:

$$dR_t^2 = [(2\alpha_t + c_t^2) + 2b_t R_t^2]dt + 2c_t R_t dW_t.$$

In finance the well-known CIR [16] interest rate model represents an important example of a squared Bessel process. As an example we simulate the generalized Bessel process modulated by Markov chain. Here we consider

$$dR_t = \frac{3}{2}(\frac{1}{R_t} - (1 - \eta_t)R_t)dt + dW_t,$$

the unobserved process $\eta = \{\eta_t, 0 \leq t\}$ is a finite state, continuous time, homogenous real valued Markov chain taking values in the set $\{\eta^1, \eta^2\}$, η^1, $\eta^2 \in \mathbb{R}$. For $t \in [0, 10]$ with $R_0 = 1$, $\eta^1 = 0.5$ and $\eta^2 = -0.5$, the intensity matrix of Markov chain $a_{12} = 0.1$, $a_{21} = 0.1$. A simulated realization of the process R_t is shown in Figure 3.5.

3.2 Quasi-Monte Carlo

Quasi-Monte Carlo simulation is the traditional Monte Carlo simulation but using quasi-random sequences instead of pseudo random numbers. In several cases, the quasi-random sequences (also called low-discrepancy sequences) improve the performance of Monte Carlo simulations, reduces computational times and /or leads to higher accuracy. In the following we simply illustrate the basic ideas behind two low-discrepancy sequences: Halton's and Sobol's sequences.

Generating Halton's low-discrepancy sequences: Halton's low-discrepancy sequences are based on a simple recipe:
Representing an integer number n in a base b, where b is a prime number:

$$n = (\cdots d_4 d_3 d_2 d_1 d_0)_b$$

Reflecting the digits and adding a radix point to obtain a number within the unit interval:

$$h = (0.d_0 d_1 d_2 d_3 d_4 \cdots)_b$$

More formally, if we represent an integer number n as

$$n = \sum_{k=0}^{m} d_k b^k,$$

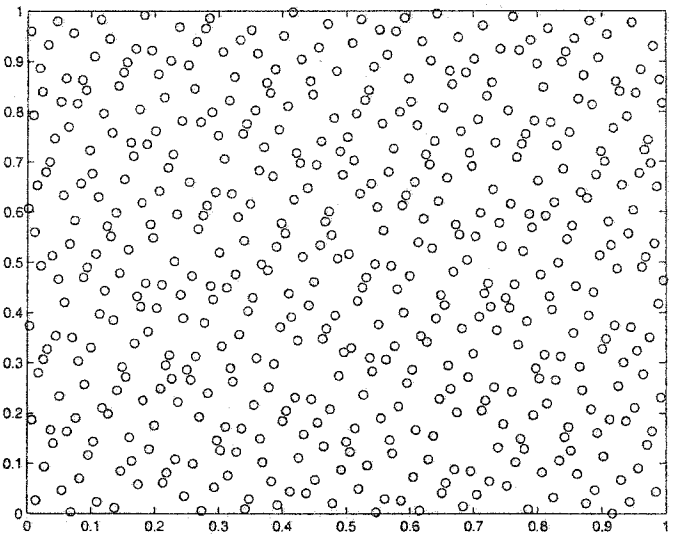

Figure 3.6: Halton low-discrepancy sequence with a given base

the nth number in the Halton's sequence with base b is

$$h(n,b) = \sum_{k=0}^{m} d_k b^{-(k+1)}.$$

Using the principles on the binary representation of numbers on a computer, it is easy to generate the nth number in a Halton's sequence with base b. (see Figure 3.6).

Generating Sobol's low-discrepancy sequences: Halton's low-discrepancy sequences are arguably the simplest, but not necessarily the best. Choosing the best sequence in practice is still an open problem, but we would like at least to take a look at a more sophisticated alternative, i.e., Sobol's sequences. For the sake of clarity, it is better to consider the generation of a one-dimensional sequence x^n in the [0,1] interval. A Sobol's sequence is generated on the basis of a set of "direction numbers" $v_1, v_2, \cdots$; we will see shortly how direction numbers are selected, but for now just think of them as numbers which are less than 1. To get the nth number in the sequence, consider

the binary representation of the integer n:

$$n = (\cdot b_3 b_2 b_1)_2.$$

The result is obtained by computing the bitwise exclusive or of the direction numbers v_i for which $b_i \neq 0$:

$$x^n = b_1 v_1 \oplus b_2 v_2 \oplus \cdots$$

If direction numbers are chosen properly, a low-discrepancy sequence will be generated(Sobol [72]). A direction number may be thought as a binary fraction:

$$v_i = (0.v_{i1} v_{i2} v_{i3} \cdots)_2,$$

or as

$$v_i = \frac{m_i}{2^i},$$

where $m_i \leq 2^i$ is an odd integer (Brandimarte [9]).

3.3 Numerical Methods for Stochastic Differential Equations

3.3.1 The Euler Scheme

The Euler approximation is a basic discrete time method to approximate an Itó process. Consider an Itó process $X = \{X(t), t_0 \leq t \leq T\}$ following the scalar stochastic differential equation

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \tag{3.3.1}$$

with $t_0 \leq t \leq T$ and the initial condition $X_{t_0} = X_0$. A discretization $t_0 = t_0 < t_1 < \cdots < t_N = T$ of the time interval $[0, T]$ may be given. Then a continuous time stochastic process $Y = \{Y(t), t_0 \leq t \leq T\}$ with the initial condition

$$Y_0 = X_0,$$

satisfying the stochastic iterative scheme

$$Y_{n+1} = Y_n + a(t_n, Y_n)(t_{n+1} - t_n) + b(t_n, Y_n)(W_{t_{n+1}} - W_{t_n}),$$

for $n = 0, 1, \cdots, N-1$, where we denote $Y_n = Y(t_n)$ is called an Euler approximation of X. The scheme is called the Euler scheme.

With the notations

$$\Delta_n = t_{n+1} - t_n, \quad \Delta W_n = W_{t_{n+1}} - W_{t_n},$$

and

$$a = a(t_n, Y_n), \quad b = b(t_n, Y_n).$$

We can write the Euler scheme in the form

$$Y_{n+1} = Y_n + a\Delta_n + b\Delta W_n,$$

for $n = 0, 1, \cdots, N-1$. In order to compute the sequence $\{Y_n, n = 0, 1, \cdots, N-1\}$ of values of the Euler approximation we have to generate the random increments ΔW_n for $n = 0, 1, \cdots, N-1$ of the Wiener process $W = \{W_t, t \geq 0\}$. These increments are independent Gaussian random variables with $E(\Delta W_n) = 0$ and $\mathrm{Var}(\Delta W_n) = \Delta_n$ and can be generated by one of the random generators described previously for independent Gaussian pseudo-random numbers, for example the Box-Muller generator.

Note that when the diffusion coefficient b is identically zero the stochastic iterative scheme reduces to the well-known deterministic Euler scheme for the ordinary differential equation $x' = a(t, x)$.

3.3.2 The Milstein Scheme

By adding to the Euler scheme the term

$$\frac{1}{2}bb'[(\Delta W_n)^2 - \Delta_n].$$

We obtain the Milstein scheme

$$Y_{n+1} = Y_n + a\Delta + b\Delta W_n + \frac{1}{2}bb'[(\Delta W_n)^2 - \Delta_n].$$

We consider the Black-Scholes model. A non-dividend paying asset, S, following geometric Brownian motion is given by the following stochastic differential equation:

$$dS_t = S_t\mu dt + S_t\sigma dW_t.$$

Where μ and σ are known constants. From Itó formula we have:

$$S_t = S_0 \exp(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t).$$

We can simulate at fixed dates:

$$S_{t_{i+1}} = S_{t_i} \exp((\mu - \frac{1}{2}\sigma^2)(t_{i+1} - t_i) + \sigma(W_{t_{i+1}} - W_{t_i})).$$

Euler Scheme:

$$S_{t_{i+1}} = S_{t_i}(1 + \mu(t_{i+1} - t_i)) + S_{t_i}\sigma(W_{t_{i+1}} - W_{t_i}).$$

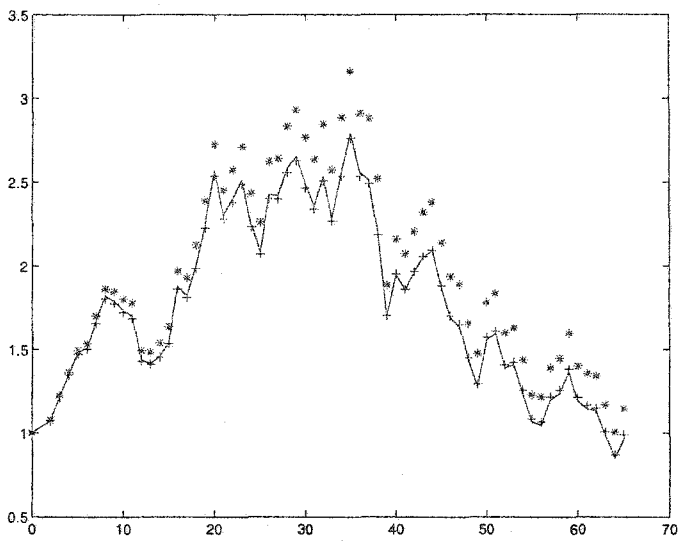

Figure 3.7: Explicit solution "-", Euler Scheme "*", Milstein Scheme "+"

Milstein Scheme:

$$\begin{aligned} S_{t_{i+1}} &= S_{t_i}(1+\mu(t_{i+1}-t_i)) + S_{t_i}\sigma(W_{t_{i+1}}-W_{t_i}) \\ &\quad +\frac{1}{2}\sigma^2 S_{t_i}((W_{t_{i+1}}-W_{t_i})^2-(t_{i+1}-t_i)). \end{aligned}$$

Figure 3.7 compares the simulation result of the explicit solution with Euler Scheme and Milstein Scheme.

3.3.3 The Generalized Discretization Scheme

In this section we extend the Euler Scheme and Milstein Scheme to diffusion with jump model.

Suppose X_t, $t \geq 0$, is a (real) process defined as the solution of a stochastic differential equation.

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t + h(t, X_{t^-})dN_t. \tag{3.3.2}$$

Here W_t, $t \geq 0$, is a standard Brownian motion and $N(t)$, $t \geq 0$ is a standard Poisson process on a probability space$(\Omega, \mathcal{F}, P)$. If X_s is known, X_t is given by

$$X_t = X_s + \int_s^t f(r, X_r)dr + \int_s^t g(r, X_r)dW_r + \int_s^t h(r, X_{r-})dN_r. \tag{3.3.3}$$

and a first approximation is to write

$$X_t \simeq X_s + f(s, X_s)(t-s) + g(s, X_s)(W_t - W_s) + h(s, X_s)(N_t - N_s). \tag{3.3.4}$$

Suppose X_r is approximated as in (3.3.4), that this approximation is substituted in (3.3.3) and that the terms are expanded as a Taylor series. Then an approximation for X_t is

$$\begin{aligned} X_s \;+\; & \int_s^t \{f(s, X_s) + \frac{\partial f}{\partial s}(s, X_s)(r-s) + \frac{\partial f}{\partial X}(s, X_s)[f(s, X_s)(r-s) \\ +\; & g(s, X_s)(W_r - W_s) + h(s, X_s)(N_r - N_s)]\}dr \\ +\; & \int_s^t \{g(s, X_s) + \frac{\partial g}{\partial s}(s, X_s)(r-s) + \frac{\partial g}{\partial X}(s, X_s)[f(s, X_s)(r-s) \\ +\; & g(s, X_s)(W_r - W_s) + h(s, X_s)(N_r - N_s)]\}dW_r \\ +\; & \int_s^t \{h(s, X_s) + \frac{\partial h}{\partial s}(s, X_s)(r-s) + \frac{\partial h}{\partial X}(s, X_s)[f(s, X_s)(r-s) \\ +\; & g(s, X_s)(W_r - W_s) + h(s, X_s)(N_{r-} - N_s)]\}dN_r. \end{aligned} \tag{3.3.5}$$

Note:

$$\int_s^t (W_r - W_s)dW_r = \frac{1}{2}((W_t - W_s)^2 - (t-s))$$

$$\int_s^t (N_{r-} - N_s)dN_r = \frac{1}{2}((N_t - N_s)^2 - (N_t - N_s))$$

$$E(W_t - W_s)^2 = t - s$$

$$E(N_t - N_s) = t - s$$

$$E(N_t - N_s)^2 = (t-s)^2 + (t-s),$$

so including only terms up to order$(t-s)$ (in expectation).

A better second order approximation (a second approximation) for X_t is

$$\begin{aligned} X_t &\simeq X_s + f(s,X_s)(t-s) + g(s,X_s)(W_t - W_s) + h(s,X_s)(N_t - N_s) \\ &+ \frac{1}{2}\frac{\partial g}{\partial X}(s,X_s)g(s,X_s)((W_t - W_s)^2 - (t-s)) \end{aligned}$$

We note in equation (3.3.2) when $h(t,X_{t^-}) = 0$, equation (3.3.3) is the same as Euler Scheme, equation (3.3.5) is the same as Milstein Scheme.
We apply the first approximation and second approximation in the following example.

We consider the Merton's ([59]) model again:

$$dS_t = S_t\mu dt + S_t\sigma dW_t + S_{t^-}d\left(\sum_{i=1}^{N(t)} Y_i\right),$$

from equation(2.2.10), we know the explicit solution is

$$S_t = S_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t)\Pi_{i=1}^{N(t)}(1+Y_i).$$

We can simulate at fixed dates:

$$S_{t_{i+1}} = S_{t_i} \exp((\mu - \frac{1}{2}\sigma^2)(t_{i+1} - t_i) + \sigma(W_{t_{i+1}} - W_{t_i}))\Pi_{j=N(t_i)+1}^{N(t_{i+1})}(1+Y_j),$$

or from jump to jump

$$S_{\tau_{j+1}^-} = S_{\tau_j} \exp((\mu - \frac{1}{2}\sigma^2)(\tau_{j+1} - \tau_j) + \sigma(W_{\tau_{j+1}} - W_{\tau_j}))$$

and

$$S_{\tau_{j+1}} = S_{\tau_{j+1}^-}(1+Y_{j+1}).$$

The first approximation:

$$\begin{aligned} S_{t_{i+1}} &= S_{t_i}(1+\mu)(t_{i+1}-t_i) + S_{t_i}\sigma(W_{t_{i+1}} - W_{t_i}) \\ &\quad + S_{t_i} \sum_{j=N(t_i)+1}^{N(t_{i+1})} Y_j. \end{aligned}$$

The second approximation:

$$\begin{aligned} S_{t_{i+1}} &= S_{t_i}(1+\mu)(t_{i+1}-t_i) + S_{t_i}\sigma(W_{t_{i+1}} - W_{t_i}) \\ &\quad + \frac{1}{2}\sigma^2 S_{t_i}((W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1}-t_i)) + S_{t_i} \sum_{j=N(t_i)+1}^{N(t_{i+1})} (Y_j + 1). \end{aligned}$$

The simulation results are compared in Figure 3.8,

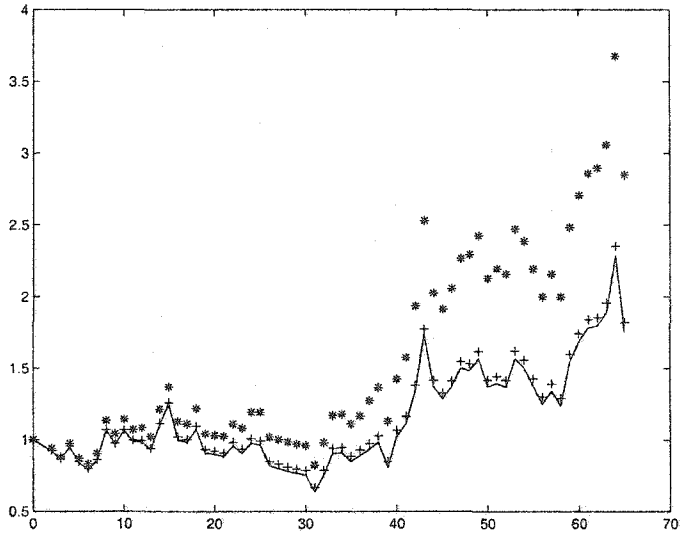

Figure 3.8: Explicit solution "-", the first approximation "*", the second approximation "+"

3.3.4 The EM Algorithm

The expectation maximization (EM) algorithm was first developed as a method for estimation in hidden Markov models (Baum and Eagon [4], Baum, Petrie, Soules and

Weiss [5]), and was later extended to a broader class of problems (Sundberg [73], Dempster, Laird and Rubin [23]). It is a general method for computing MLEs in statistical models in which there exist unobservable random variables. Suppose that we can observe some variable Y, but that there exist additional variable, X, that we can not observe. This problem is nicely discussed by Dembo and Zeitouni [21], [22] where observations are diffusion case and also are used by Ryden [67], [68] to estimate the parameters of a MMPP (Markov-modulated Poisson process), Elliott et. al use these methods in many filtering literature. We briefly review the EM algorithm below and will use this method to re-estimate parameters in Chapter 4 and Chapter 5.

Suppose $\{P_\theta, \theta \in \Theta\}$ is a family of probability measures on $(\Omega, \mathcal{F})$, all absolutely continuous with respect to a fixed probability measure P. The likelihood function of the parameter θ based on $\mathcal{Y}_t$ is $L(\theta) = E[(dP_\theta/dP_{\theta^*})|\mathcal{Y}_t]$. The MLE is difficult to compute directly. In such cases, the EM algorithm is a convenient iterative numerical method for computing the MLE. Each iteration of the EM algorithm consists of four steps.

- Step 1: Set $p = 0$, and choose $\hat{\theta}_0$.
- Step 2: (E-step) Set $\theta^* = \hat{\theta}_p$, and compute $Q(\cdot, \theta^*)$, where
$$Q(\theta, \theta^*) = E_{\theta^*}[\log \frac{dP_\theta}{dP_{\theta^*}} \mid \mathcal{Y}_t].$$
- Step 3: (M-step) Find
$$\hat{\theta}_{p+1} \in \arg\max_{\theta \in \Theta} Q(\theta, \theta^*).$$
- Step 4: Replace p by $p + 1$ and repeat beginning with step 2 until a stoping criterion is satisfied.

The sequence generated $\{\hat{\theta}_p, p \geq 0\}$ gives nondecreasing values of the likelihood function. Indeed, it follows from Jensen's inequality, that $\log L(\hat{\theta}_{p+1}) - \log L(\hat{\theta}_p) \geq Q(\hat{\theta}_{p+1}, \hat{\theta}_p)$, with equality if and only if $\hat{\theta}_{p+1} = \hat{\theta}_p$. We call $Q(\theta, \theta^*)$ a conditional log-likelihood.

Chapter 4

Hidden Markov Chain Filtering for Diffusion-Jump Model

4.1 Introduction

The powerful tools of stochastic calculus are finding their way into many branches, especially in filtering and finance. They have enabled analysis of more complicated models than could be handled earlier, an aspect of this development is the growing use of diffusion with jumps processes. Since Merton[59] proposed the general stock price model, diffusion with jumps processes such as defaultable term structure models and energy market with spikes have been widely used in financial literature. It is a challenge to estimate the parameters of the models and latent variables. Motivated by this aim, we propose the hidden Markov filtering for diffusion with jumps process.

We obtain recursive filters for the latent states of the Markov chain, the number of jumps from one state to another, and the occupation time of the Markov chain in any state. The filter equations are finite-dimensional and in closed form. Using the EM algorithm, we obtain the estimators for parameters and jump intensity rate of the model.

This Chapter is organized as follows: In Section 4.2, we present the problem formulation. In Section 4.3, the change of probability measure is given. The derivation of Girsanov's transformation for diffusion with jump process in detail, it is also an excellent example of the generalized Itó formula for semi-martingales. In Section 4.4, Hidden Markov chain filters are given. In Section 4.5, we extend the EM algorithm for diffusion with jumps process and revise the model parameters and jump intensity rates. Section 4.6 is conclusion. In our work, the main contribution is in two ways: we develop the filtering techniques to diffusion with jumps processes and derive finite-dimensional filters for various statistics including the Markov chain state, the number of jumps between states, and the occupation time in a state. The filters that we derived compute all statistics required to implement the E-step of EM algorithm; and on the other hand we extend the EM algorithm.

4.2 Problem Formulation

All random variables are defined on the probability space $(\Omega, \mathcal{F}, P)$. Let X_t, $t \geq 0$ be a finite state, continuous time, homogeneous Markov chain. Without loss of generality, X_t takes values in the set $\Sigma = \{e_1, e_2, \cdots, e_s\}$ of unit (column) vectors, $e_i = (0, \ldots, 1, \ldots, 0)'$ of R^S. Suppose the time homogeneous hidden Markov chain X has intensity-matrix A, $A = (a_{ij})$, $1 \leq i, j \leq S$, where the intensity for a jump from e^i into e^j is denoted by $a_{ij} \geq 0$, $i \neq j$. That is, defining $p_t^i = P(X_t = e_i)$, $1 \leq i \leq S$, the probability distribution $p_t = (p_t^1, p_t^2, \cdots, p_t^S)'$ satisfies the forward Kolmogorov equation $dp_t/dt = Ap_t$. Also note that $\sum_{i=1}^{S} a_{ij} = 0$, for $1 \leq j \leq s$. We assume that

$E\{X_0\}$ is known.

The process X_t is not observed directly, but rather we suppose there is a (scalar) observation process given by

$$y_t = \int_0^t g(X_r)dr + \xi W_t + N_t. \tag{4.2.1}$$

(the extension to vector process Y is straightforward). $W = \{W_t : 0 \leq t\}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, P)$, which is independent of X_t. ξ is a constant. $N = \{N_t : 0 \leq t\}$ is a doubly stochastic Poisson process (Cox process) with intensity rate λ_t that may depend on X_t on a given probability space $(\Omega, \mathcal{F}, P)$. N_t denotes the number of events that occurs during the interval $[0, t]$. Because X_t takes values in Σ, any real function $h(X_t)$ can be given by a vector $h = (h_1, \ldots, h_S) \in R^S$, and $h(X_t) = \langle h, X_t \rangle$, where $\langle . \rangle$ denotes the scalar product in R^S. Consequently, there are vectors $g = (g_1, g_2, \cdots, g_S)'$ and $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_S)'$, so that $g(X_t) = \langle g, X_t \rangle$ and $\lambda(X_t) = \langle \lambda, X_t \rangle$. The observation process has the form:

$$y_t = \int_0^t \langle g, X_r \rangle dr + \xi W_t + N_t. \tag{4.2.2}$$

Defining the increasing families of $\sigma-$subalgebras: $\mathcal{F}_t = \mathcal{F}_t^x \vee \mathcal{Y}_t$, where $\mathcal{F}_t^x = \sigma\{X_s, : s \leq t\}$ and $\mathcal{Y}_t = \sigma\{y_s : s \leq t\}$, each contains all P-null subsets in $\mathcal{F}$. For simplicity, we make further assumptions on the independence of Poisson process and Markov processes.

Aim: Given $\mathcal{Y}_t$, $t \geq 0$, estimate the statistics of the model and obtain filtered estimates of:

- The Markov chain state X_t, namely $\hat{X}_t = E\{X_t | \mathcal{Y}_t\}$.

- The number of jumps J_t^{ij} of Markov chain from state e_i to e_j in the time interval $[0,t]$:

$$J_t^{ij} = \int_0^t \langle X_{r-}, e_i \rangle \langle e_j, dX_r \rangle, \qquad 1 \leq i,j \leq S. \tag{4.2.3}$$

- The occupation time O_t^i of the Markov chain in the state e_i in the interval $[0,t]$:

$$O_t^i = \int_0^t \langle X_r, e_i \rangle dr, \qquad 1 \leq i \leq S. \tag{4.2.4}$$

- Further, we write:

$$G_t^i = \int_0^t \langle X_r, e_i \rangle dy_r, \qquad 1 \leq i \leq S. \tag{4.2.5}$$

First, we give a preliminary result which will be used in next section.

It is straightforward to show that the semi-martingale representation of X_t is

$$X_t = X_0 + \int_0^t AX_r dr + M_t, \tag{4.2.6}$$

where M_t is a (vector) $\mathcal{F}_t$-martingale under P (Elliott[30]).

Note $\int_0^t AX_r dr = \int_0^t AX_{r-} dr$ because $X_r(w) = X_{r-}(w)$ a.s. except for a countable number of values of r. Similar identifications will be made.

We give the semi-martingale decomposition of J_t^{ij}, using:

$$\begin{aligned} J_t^{ij} &= \int_0^t \langle X_{r-}, e_i \rangle \langle e_j, AX_{r-} \rangle dr + \int_0^t \langle X_{r-}, e_i \rangle \langle e_j, dM_r \rangle \\ &= \int_0^t \langle X_{r-}, e_i \rangle \langle e_j, AX_{r-} \rangle dr + M_t^{ij}, \end{aligned} \tag{4.2.7}$$

where

$$M_t^{ij} = \int_0^t \langle X_{r-}, e_i \rangle \langle e_j, dM_r \rangle.$$

Note the integrand $\langle X_r, e_i\rangle e_j$ is predictable, so M_t^{ij} is a martingale.

Now $\langle X_r-, e_i\rangle\langle e_j, AX_r-\rangle = \langle X_r-, e_i\rangle a_{ji}$, so

$$J_t^{ij} = \int_0^t \langle X_r-, e_i\rangle a_{ji} dr + M_t^{ij}, \quad 1 \leq i, j \leq s. \tag{4.2.8}$$

We also have the decomposition of G_t^i,

$$\begin{aligned} G_t^i &= \int_0^t \langle X_r, e_i\rangle dy_r \\ &= \int_0^t g_i\langle X_r, e_i\rangle dr + \int_0^t \xi\langle X_r, e_i\rangle dW_r + \int_0^t \langle X_r, e_i\rangle dN_r \\ &= G_{1t}^i + G_{2t}^i. \end{aligned}$$

where

$$\begin{aligned} G_{1t}^i &= \int_0^t g_i\langle X_r, e_i\rangle dr + \int_0^t \xi\langle X_r, e_i\rangle dW_r, \\ G_{2t}^i &= \int_0^t \langle X_r, e_i\rangle dN_r. \end{aligned}$$

More succinctly

$$\begin{aligned} G_{1t}^i &= \int_0^t \langle X_r, e_i\rangle dy_r^c, \\ G_{2t}^i &= \int_0^t \langle X_r, e_i\rangle dy_r^d, \end{aligned}$$

where

$$\begin{aligned} y_t^c &= \int_0^t \langle g, X_r\rangle dr + \xi W_t, \\ y_t^d &= N_t. \end{aligned}$$

4.3 Change of Probability Measure

For our purpose, we want to use the so-called measure transformation approach or reference-probability method. Assuming that X_t and Y_t are defined on a probability

space $\{\Omega, \mathcal{F}, P\}$, the aim of the present section is therefore to construct a new probability measure P_0on $\{\Omega, \mathcal{F}_t\}$ that is mutually absolutely continuous with respect to P and such that under P_0 the process X_t is a finite state Markov chain with transition intensity matrix family A as before and the process $\tilde{W}_t = \xi^{-1}(y_t - N_t)$ is a standard Brownian motion, and $\tilde{N}_t = y_t - \int_0^t \langle g, X_r \rangle dr - \xi W_t$ is a standard Poisson process, X_t and Y_t are independent.

Define the probability measure P_0 such that the $\mathcal{F}_t$ restriction of the Radon-Nikodym derivative of P with respect to P_0 is:

$$\begin{aligned}\Lambda_t = \frac{dP}{dP_0}\,|_{\mathcal{F}_t} &= \exp\{\int_0^t \xi^{-1}\langle g, X_r\rangle d\tilde{W}_r - \frac{1}{2}\int_0^t \xi^{-2}\langle g, X_r\rangle^2 dr \\ &\quad + \int_0^t \log\langle \lambda, X_r\rangle d\tilde{N}_r + \int_0^t (1 - \langle \lambda, X_r\rangle) dr\}. \qquad (4.3.1)\end{aligned}$$

Note that: $\Lambda_t = \Lambda_t^c * \Lambda_t^d$, where

$$\begin{aligned}\Lambda_t^c &= \exp\{\int_0^t \xi^{-1}\langle g, X_r\rangle d\tilde{W}_r - \frac{1}{2}\int_0^t \xi^{-2}\langle g, X_r\rangle^2 dr\}, \\ \Lambda_t^d &= \exp\{\int_0^t \log\langle \lambda, X_r\rangle d\tilde{N}_r + \int_0^t (1 - \langle \lambda, X_r\rangle) dr\}.\end{aligned}$$

Denoting by $\bar{E}$ the expectation with respect to the measure P_0. If ϕ_t is an $\mathcal{F}_t$ adapted integrable process, then an abstract version of Bayes theorem states (see Chapter 2) that:

$$\hat{\phi}_t = E\{\phi_t | \mathcal{Y}_t\} = \frac{\bar{E}(\Lambda_t \phi_t | \mathcal{Y}_t)}{\bar{E}(\Lambda_t | \mathcal{Y}_t)}.$$

$\bar{E}(\cdot)$ is called the unnormalized conditional measure under P_0 of ϕ_t given $\mathcal{Y}_t$.

We have the following result hold:

Theorem 4.3.1. *(1) Λ_t^c, and Λ_t^d are $(\mathcal{F}_t, P_0)$ martingales.*

(2) Λ_t *is the unique solution to the integral equation :*

$$\Lambda_t = 1 + \int_0^t \xi^{-1}\Lambda_r\langle g, X_r\rangle d\tilde{W}_r + \int_0^t \Lambda_{r-}(\langle\lambda, X_r\rangle - 1)d(\tilde{N}_r - r). \tag{4.3.2}$$

(3) $\tilde{W}_t = \int_0^t \xi^{-1}\langle g, X_r\rangle dr + W_t$ *is an* $(\mathcal{F}_t, P_0)$ *- standard Brownian motion, and* $\tilde{N}_t$ *is an* $(\mathcal{F}_t, P_0)$ *- standard Poisson process.*

(4) $\bar{E}\{e^{iv(y_t - y_r)}|\mathcal{F}_r\} = \exp[(e^{iv} - 1 - \frac{\xi^{-2}}{2}v^2)(t-r)]$.

For notational convenience, define: $K_t = (\Lambda_t)^{-1}$, $K_t^c = (\Lambda_t^c)^{-1}$, $K_t^d = (\Lambda_t^d)^{-1}$. In fact, we have:

$$\begin{aligned} K_t = \frac{dP_0}{dP}\,|_{\mathcal{F}_t} &= \exp\{-\int_0^t \xi^{-1}\langle g, X_r\rangle dW_r - \frac{1}{2}\int_0^t \xi^{-2}\langle g, X_r\rangle^2 dr \\ &\quad - \int_0^t \log\langle\lambda, X_r\rangle dN_r - \int_0^t (1 - \langle\lambda, X_r\rangle)dr\}. \end{aligned} \tag{4.3.3}$$

Proof. (1) We have:

$$\begin{aligned} dK_t^c &= -\xi^{-1}\langle g, X_t\rangle K_t^c dW_t, \\ dK_t^d &= K_{t-}^d \frac{\langle\lambda, X_t\rangle - 1}{\langle\lambda, X_t\rangle}(dN_t - \langle\lambda, X_t\rangle dt). \end{aligned}$$

The derivation of dK_t^d, namely, $d(\Lambda_t^d)^{-1}$ is given below.

We take $x_t = \Lambda_t^d$, $f(\cdot) = (\cdot)^{-1}$, using the generalized Itó formula (2.2.1), this rule states:

$$\begin{aligned} f(x_t) &= f(x_0) + \int_0^t f'(x_{r-})dx_r + \frac{1}{2}\int_0^t f''(x_{r-})d[x,x]_r^c \\ &\quad + \sum_{0<r\le t}\{f(x_r) - f(x_{r-}) - f'(x_{r-})\Delta x_r\}. \end{aligned} \tag{4.3.4}$$

Now,

$$\begin{aligned} &\sum_{0<r\le t}\{f(x_r) - f(x_{r-})\} = \sum_{0<r\le t}(\Lambda_{r-}^d)^{-1}(\frac{1}{\langle\lambda, X_r\rangle} - 1)\Delta N_r \\ &= \sum_{0<r\le t} -(\Lambda_{r-}^d)^{-1}\frac{(\langle\lambda, X_r\rangle - 1)}{\langle\lambda, X_r\rangle}\Delta N_r. \end{aligned}$$

The remaining term in the sum over jump events of (4.3.4) is $f'(x_r)\Delta x_r$, this term is $(\Lambda^d_{r-})^{-1}(\langle\lambda, X_r\rangle - 1)\Delta N_r$. Hence the sum over jump events in (4.3.4) is

$$\sum_{0<r\leq t}\{-(\Lambda^d_{r-})^{-1}\frac{(\langle\lambda, X_r\rangle - 1)}{\langle\lambda, X_r\rangle}\Delta N_r + (\Lambda^d_{r-})^{-1}(\langle\lambda, X_r\rangle - 1)\Delta N_r\}$$
$$= \int_0^t (\Lambda^d_{r-})^{-1}\frac{(\langle\lambda, X_r\rangle - 1)}{\langle\lambda, X_r\rangle}dN_r + \int_0^t (\Lambda^d_{r-})^{-1}(\langle\lambda, X_r\rangle - 1)dN_r.$$

For other terms in (4.3.4), we have the following:

$$f(x_0) = 1,$$
$$\int_0^t f'(x_r)dx_r = -\int_0^t (\Lambda^d_{r-})^{-1}(\langle\lambda, X_r\rangle - 1)(dN_r - dr),$$
$$\frac{1}{2}\int_0^t f''(x_r)d[x^c, x^c]_r = 0.$$

Collecting all terms evaluated according to (4.3.4) and rearrangement, we get:

$$d(\Lambda^d_t)^{-1} = (\Lambda^d_{t-})^{-1}\frac{\langle\lambda, X_t\rangle - 1}{\langle\lambda, X_t\rangle}(dN_t - \langle\lambda, X_t\rangle dt).$$

It is obvious that K^c_t and K^d_t are $(\mathcal{F}_t, P)$ martingales. It is easy to prove that Λ^c_t and Λ^d_t are martingales and also independent on $(\mathcal{F}_t, P_0)$.

In order to prove (2), we use Corollary 2.2.3, this rule states:

$$X_tY_t = X_0Y_0 + \int_0^t X_{r-}dY_r + \int_0^t Y_{r-}dX_r + [X, Y]_t.$$

Now,

$$\Lambda_t = \Lambda^c_t * \Lambda^d_t,$$

so,

$$\Lambda_t = \Lambda_0 + \int_0^t \Lambda^d_{r-}d\Lambda^c_r + \int_0^t \Lambda^c_r d\Lambda^d_r + [\Lambda^c, \Lambda^d]_t.$$

However,

$$[\Lambda^c, \Lambda^d]_t = 0,$$

$$d\Lambda_r^c = \xi^{-1}\Lambda_r^c \langle g, X_r\rangle d\tilde{W}_r,$$

$$d\Lambda_r^d = \Lambda_{r-}^d(\langle \lambda, X_r\rangle - 1)d(\tilde{N}_r - r).$$

Therefore,

$$\begin{aligned}\Lambda_t &= 1 + \int_0^t \Lambda_{r-}^d \Lambda_r^c \xi^{-1}\langle g, X_r\rangle d\tilde{W}_r \\ &\quad + \int_0^t \Lambda_r^c \Lambda_{r-}^d(\langle \lambda, X_r\rangle - 1)d(\tilde{N}_r - r),\end{aligned} \tag{4.3.5}$$

namely,

$$\Lambda_t = 1 + \int_0^t \Lambda_r \xi^{-1}\langle g, X_r\rangle d\tilde{W}_r + \int_0^t \Lambda_{r-}(\langle \lambda, X_r\rangle - 1)d(\tilde{N}_r - r).$$

(3) By Bayes' formula:

$$\begin{aligned}\tilde{E}[\tilde{W}_t|\mathcal{F}_s] &= \frac{E[K_t\tilde{W}_t|\mathcal{F}_s]}{E[K_t|\mathcal{F}_s]} = \frac{E[K_t^d K_t^c \tilde{W}_t|\mathcal{F}_s]}{E[K_t^d K_t^c|\mathcal{F}_s]} \\ &= (K_s^c)^{-1}E[K_t^c\tilde{W}_t|\mathcal{F}_s],\end{aligned}$$

where

$$\tilde{W}_t = \int_0^t \xi^{-1}\langle g, X_r\rangle dr + W_t.$$

Clearly $\tilde{W}_t$ is a P_0- martingale if and only if, $K_t^c\tilde{W}_t$ is a P- martingale. Now

$$K_t^c\tilde{W}_t = \int_0^t K_r^c d\tilde{W}_r + \int_0^t \tilde{W}_r dK_r^c + [K^c, \tilde{W}]_t.$$

However,

$$[K^c, \tilde{W}]_t = \int_0^t K_r^c \xi^{-1}\langle g, X_r\rangle dr.$$

Therefore,

$$K_t^c \tilde{W}_t = \int_0^t K_r^c dW_r + \int_0^t \tilde{W}_r K_r^c dW_r.$$

Because W_t is a martingale under P, by the characterization of Brownian Motion due to Lévy, the result follows. By the same argument we have:

$$\begin{aligned} K_t^d \tilde{N}_t &= \int_0^t K_{r-}^d d\tilde{N}_r + \int_0^t \tilde{N}_{r-} dK_r^d + [K^d, \tilde{N}]_t \\ [K^d, \tilde{N}]_t &= \sum_{0<r\leq t} \Delta K_r^d \Delta \tilde{N}_r = \int_0^t -K_r^d \frac{\langle \lambda, X_r \rangle - 1}{\langle \lambda, X_r \rangle} dN_r. \end{aligned}$$

There,

$$\begin{aligned} K_t^d(\tilde{N}_t - t) &= \int_0^t K_{r-}^d \frac{1}{\langle \lambda, X_r \rangle}(dN_r - \langle \lambda, X_r \rangle dr) \\ &\quad + \int_0^t \tilde{N}_{r-} K_{r-}^d \frac{\langle \lambda, X_r \rangle - 1}{\langle \lambda, X_r \rangle}(dN_r - \langle \lambda, X_r \rangle dr). \end{aligned}$$

Because $N_t - \int_0^t \langle \lambda, X_r \rangle dr$ is a martingale under P, by the characterization of stochastic Poisson Process due to S. Watanabe (Brémaud [8]), the result follows.

(4) We have:

$$\begin{aligned} y_t - y_s &= \int_s^t \langle g, X_r \rangle dr + \xi(W_t - W_s) + N_t - N_s \\ &= \xi^{-1}(\tilde{W}_t - \tilde{W}_s) + (\tilde{N}_t - \tilde{N}_s). \end{aligned}$$

Where $\tilde{W}_t$, $\tilde{N}_t$ standard Brownian motion and standard Poisson process under P_0, use Lévy-Khintchine formula, (4) follows.

□

It is obvious that we also have the result: P_0 is a probability measure, $P_0 \sim P$, under P_0 the process Y_t has independent increments, and X_t, Y_t are independent. From (3) we have:

$$\begin{aligned} \bar{E}\{e^{iv(y_t - y_r)} | \mathcal{Y}_r\} &= \bar{E}\{\bar{E}\{e^{iv(y_t - y_r)} | \mathcal{F}_r\} | \mathcal{Y}_r\} \\ &= \exp[(e^{iv} - 1 - \frac{\xi^2}{2} v^2)(t - r)], \end{aligned}$$

which shows that Y_t is an independent increment process. Analogously:

$$\begin{aligned}\bar{E}\{e^{iv(y_t-y_r)}|\mathcal{F}_r^x\} &= \bar{E}\{\bar{E}e^{iv(y_t-y_r)}|\mathcal{F}_r\}|\mathcal{F}_r^x\} \\ &= \exp[(e^{iv}-1-\frac{\xi^{-2}}{2}v^2)(t-r)],\end{aligned}$$

which shows that the increment distributions, and therefore all joint distributions of Y_t, do not depend on $\{X_t, 0 \leq t \leq T\}$.

4.4 Hidden Markov Chain Filters

With the measure P_0 introduced in the previous section, it follows the Kallianpur-Striebel-Bayes formula that the optimal filter can be given the following representation:

$$E(f(X_t))|\mathcal{Y}_t) = \frac{\bar{E}(\Lambda_t f(X_t)|\mathcal{Y}_t)}{\bar{E}(\Lambda_t|\mathcal{Y}_t)} \tag{4.4.1}$$

For notational convenience, we define $\sigma(f(X_t)) = \bar{E}(\Lambda_t f(X_t)|\mathcal{Y}_t)$. Then, $\hat{f}(X_t)$ can be re-expressed as

$$\hat{f}(X_t) = \frac{\sigma(f(X_t))}{\hat{\Lambda}_t},$$

where $\hat{f}(X_t) \triangleq E(f(X_t)|\mathcal{Y}_t)$, $\hat{\Lambda}_t \triangleq \sigma(1) = \bar{E}(\Lambda_t|\mathcal{Y}_t)$.

Since under P_0 the process X_t and Y_t are independent, the condition in the right hand side of (4.4.1) only fixes the trajectory of the process Y_t while the expectation is an ordinary expectation over the process X_t. It proves convenient to work with the unnormalized as it satisfies an equation less complicated than its corresponding normalized version.

Theorem 4.4.1. *The unnormalised filter $\sigma(X)$ is the solution of the following vector stochastic equation:*

$$\begin{aligned}\sigma(X_t) &= \sigma(X_0) + \int_0^t A\sigma(X_r)dr + \int_0^t B\sigma(X_r)d\tilde{W}_r \\ &\quad + \int_0^t G\sigma(X_{r-})d(\tilde{N}_r - r), \end{aligned} \tag{4.4.2}$$

where, $B = \xi^{-1}diag\{\langle g, e_i\rangle\}$, $G = diag\{\langle \lambda, e_i\rangle - 1\}$.

Proof. In order to obtain the filtered estimate, we use general form of Itó formula:

$$\begin{aligned}\Lambda_t X_t &= X_0\Lambda_0 + \int_0^t X_{r-}d\Lambda_r + \int_0^t \Lambda_{r-}dX_r + [\Lambda, M]_t \\ &= X_0\Lambda_0 + \int_0^t \xi^{-1}X_{r-}\Lambda_r\langle g, X_r\rangle d\tilde{W}_r + \int_0^t \Lambda_{r-}AX_r dr \\ &\quad + \int_0^t X_{r-}\Lambda_{r-}(\langle \lambda, X_r\rangle - 1)d(\tilde{N}_r - r) + \int_0^t \Lambda_{r-}dM_r + [\Lambda, M]_t,\end{aligned}$$

for $t \geq 0$, conditioning each side of the equation above on $\mathcal{Y}_t$, we obtain that:

$$\begin{aligned}\sigma(X_t) &= \sigma(X_0) + \int_0^t A\sigma(X_r)dr - \int_0^t \sigma(X_r)d(\tilde{N}_r - r) \\ &\quad + \int_0^t \bar{E}(X_{r-}\Lambda_{r-}\langle \lambda, X_r\rangle|\mathcal{Y}_t)d(\tilde{N}_r - r) \\ &\quad + \int_0^t \xi^{-2}\bar{E}(X_r\Lambda_r\langle g, X_r\rangle|\mathcal{Y}_t)d\tilde{W}_r.\end{aligned}$$

Note that: $\sum_{i=1}^S \langle X_r, e_i\rangle = 1$, we have:

$$\begin{aligned}\bar{E}(X_r\Lambda_r\langle g, X_r\rangle|\mathcal{Y}_t) &= \bar{E}(X_r\Lambda\langle g, X_r\rangle \sum_{i=1}^S \langle X_r, e_i\rangle|\mathcal{Y}_t) \\ &= \sum_{i=1}^S \bar{E}(\langle X_r^2\Lambda_r\langle g, X_r\rangle, e_i\rangle|\mathcal{Y}_t) \\ &= \sum_{i=1}^S e_i\langle g, e_i\rangle\langle \sigma(X_r), e_i\rangle.\end{aligned}$$

Similarly, we have:

$$\bar{E}(X_r\Lambda_r\langle \lambda, X_r\rangle|\mathcal{Y}_r) = \sum_{i=1}^S e_i\langle \lambda, e_i\rangle\langle \sigma(X_r), e_i\rangle,$$

which implies (4.4.2).

□

4.5 Revising the Parameters

In this section, we will use the EM algorithm to re-estimate the hidden Markov chain transition matrix and Poisson process intensity rate $\lambda(X_t)$. The E-step can be implemented using filters, our work is a direct extension of the papers mentioned in Chapter 3 (Ryden [67], [68], Dembo and Zeitouni [21], [22]).

Suppose $\{P_\theta, \theta \in \Theta\}$ is a family of probability measures on $(\Omega, \mathcal{F})$, all absolutely continuous with respect to a fixed probability measure P. The likelihood function of the parameter θ based on $\mathcal{Y}_t$ is $L(\theta) = E[(dP_\theta/dP_{\theta'})|\mathcal{Y}_t]$.

Our model is determined by the set of parameters:

$$\theta = \{a_{ji}, g_i, \lambda_i, 1 \leq i, j \leq S\}.$$

Further $a_{ji} > 0$, for $i \neq j$, $\sum_{j=1}^{S} a_{ji} = 0$, so that there is no need to estimate the elements a_{ii}, $i = 1, \ldots, S$, if the others are already estimated. Suppose our model is given by such a set of parameters and we wish to determine a new set:

$$\hat{\theta} = \{\hat{a}_{ji}, \hat{g}_i, \hat{\lambda}_i, 1 \leq i, j \leq S\},$$

which maximizes the conditional log-likelihoods. To change all the a_{ji} to $\hat{a}_{ji}$, to

change the g_i to $\hat{g}_i$ and λ_i to $\hat{\lambda}_i$, we use (4.3.3) and define:

$$\begin{aligned}\frac{dP_\theta}{dP_{\theta'}}|_{\mathcal{F}_t} &= \prod_{i,j=1,i\neq j}^{S} L_t^{ij} \exp\big(-\int_0^t \xi^{-1}\langle g-g', X_r\rangle d\tilde{W}_r \\ &\quad -\frac{1}{2}\int_0^t \xi^{-2}\langle g, X_r\rangle^2 dr - \frac{1}{2}\int_0^t \xi^{-2}\langle g', X_r\rangle^2 dr \\ &\quad -\int_0^t (\log\langle\lambda, X_r\rangle - \log\langle\lambda', X_r\rangle) d\tilde{N}_r + \int_0^t \langle\lambda-\lambda', X_r\rangle dr\big).\end{aligned}$$

Now with:

$$L_t^{ij} = \Big(\frac{a_{ji}}{a'_{ji}}\Big)^{J_t^{ij}} \exp\big(\int_0^t (a'_{ji} - a_{ji})\langle X_r, e_i\rangle dr\big). \tag{4.5.1}$$

Then we compute the quantity (E-step):

$$Q(\theta, \theta') = E\Big(\log \frac{dP_\theta}{dP_{\theta'}}|_{\mathcal{F}_t}\Big). \tag{4.5.2}$$

Using the fact:

$$\langle g, X_r\rangle = \sum_i^S g_i\langle X_r, e_i\rangle, \qquad \langle\lambda, X_r\rangle = \sum_i^S \lambda_i\langle X_r, e_i\rangle. \tag{4.5.3}$$

Further,

$$\begin{aligned}\langle g, X_r\rangle^2 &= \sum_{i=1}^S g_i^2\langle X_r, e_i\rangle, \\ \log\langle\lambda, X_r\rangle &= \sum_{i=1}^S \log\lambda_i\langle X_r, e_i\rangle.\end{aligned}$$

We have:

$$\begin{aligned}
\log\left(\frac{dP_\theta}{dP_{\theta'}}|_{\mathcal{F}_t}\right) &= \sum_{i,j=1,i\neq j}^{S} \log\left(L_t^{ij}\right) \int_0^t \xi^{-1}\langle g-g', X_r\rangle d\tilde{W}_r \\
&\quad -\frac{1}{2}\int_0^t \xi^{-2}\langle g, X_r\rangle^2 dr - \frac{1}{2}\int_0^t \xi^{-2}\langle g', X_r\rangle^2 dr \\
&\quad -\int_0^t (\log\langle\lambda, X_r\rangle - \log\langle\lambda', X_r\rangle) d\tilde{N}_r - \int_0^t \langle\lambda-\lambda', X_r\rangle dr \\
&= \sum_{i,j=1,i\neq j}^{S} \left(J_t^{ij}(\log a_{ij} - \log a'_{ij}) + \int_0^t (a'_{ji} - a_{ji})\langle X_r, e_i\rangle dr\right) \\
&\quad -\int_0^t \xi^{-1}\langle g-g', X_r\rangle d\tilde{W}_r - \frac{1}{2}\int_0^t \xi^{-2}(\langle g, X_r\rangle^2 - \langle g', X_r\rangle^2) dr \\
&\quad -\int_0^t (\log\langle\lambda, X_r\rangle - \log\langle\lambda', X_r\rangle) d\tilde{N}_r + \int_0^t \langle\lambda-\lambda', X_r\rangle dr.
\end{aligned} \tag{4.5.4}$$

Therefore, taking the conditional expectation on $\mathcal{Y}_t$ of both sides (4.5.4), we obtain:

$$\begin{aligned}
Q(\theta,\theta') &= \sum_{i,j=1,i\neq j}^{S} \left(\hat{J}_t^{ij}\log a_{ji} - a_{ji}\hat{O}_t^i\right) + \sum_{i=1}^{S}\left(\xi^{-1}g_i\hat{G}_{1t}^i - \frac{\xi^{-2}}{2}g_i^2\hat{O}_t^i\right) \\
&\quad + \sum_{i=1}^{S}\left(\hat{G}_{2t}^i\log\lambda_i - \lambda_i\hat{O}_t^i\right) + R(\theta'),
\end{aligned} \tag{4.5.5}$$

where the term $R(\theta')$ does not involve the parameter θ, so the E-step consists of computing $\hat{J}_t^{ij}$, $\hat{O}_t$ and $\hat{G}_t$.

For the M-step, set the partial derivatives with respect to a_{ji}, g_i and λ_i to zero, this yields the estimates $\hat{\theta}_{p+1}$ as

$$\begin{aligned}
\hat{a}_{ji} &= \hat{J}_t^{ij}/\hat{O}_t^i = \sigma(J_t^{ij})/\sigma(O_t^i), \\
\hat{g}_i &= \xi^{-1}\hat{G}_{1t}^i/\hat{O}_t^i = \xi^{-1}\sigma(G_{1t}^i)/\sigma(O_t^i), \\
\hat{\lambda}_i &= \hat{G}_{2t}^i/\hat{O}_t^i = \sigma(G_{2t}^i)/\sigma(O_t^i).
\end{aligned}$$

We now derive filters for J_t^{ij}, O_t^i, G_{1t}^i and G_{2t}^i which are needed to compute $a_{ij}^{'}$, $g_i^{'}$, $\lambda_i^{'}$ and give $\sigma(J_t^{ij}X_t)$, $\sigma(O_t^iX_t)$, $\sigma(G_{1t}^iX_t)$ and $\sigma(G_{2t}^iX_t)$ to satisfy the stochastic differential equations.

If $\sigma(J_t^{ij}X_t)$, for example, is determined, then: $\sigma(J_t^{ij}) = \sum_{i=1}^{S}\langle\sigma(J_t^{ij}X_t), e_i\rangle$. In fact,

$$\begin{aligned}\sigma(J_t^{ij}) &= \bar{E}(J_t^{ij}\Lambda_t|\mathcal{Y}_t) = \bar{E}(J_t^{ij}\Lambda_t\sum_{k=1}^{S}\langle X_t, e_k\rangle|\mathcal{Y}_t)\\ &= \sum_{k=1}^{S}\langle\sigma(J_t^{ij}X_t), e_k\rangle.\end{aligned}$$

Theorem 4.5.1. *For $0 \le t$ and $1 \le i, j \le S$, $i \ne j$*

$$\begin{aligned}\sigma(J_t^{ij}X_t) &= \int_0^t A\sigma(J_r^{ij}X_r)dr + \int_0^t B\sigma(J_r^{ij}X_r)d\tilde{W}_r\\ &\quad + \int_0^t\langle\sigma(X_r), e_i\rangle a_{ji}e_j dr + \int_0^t G\sigma(J_r^{ij}X_r)d(\tilde{N}_r - r). \qquad (4.5.6)\end{aligned}$$

The semi-martingale representation of the Markov chain X is given by (4.2.6) and the semi-martingale decomposition of J_t^{ij} is given by (4.2.7), we have:

$$\begin{aligned}J_t^{ij}X_t &= \int_0^t X_r\langle X_{r-}, e_i\rangle a_{ij}dr + \int_0^t X_{r-}dM_r^{ij} + \int_0^t AJ_r^{ij}dr\\ &\quad + \int_0^t J_{r-}^{ij}dM_r + \sum_{0<r\le t}(e_j - e_i)\langle X_{r-}, e_i\rangle\langle X_r, e_j\rangle\\ &= \int_0^t\langle X_r, e_i\rangle a_{ij}e_j dr + \int_0^t AJ_r^{ij}X_r dr + \int_0^t\bar{\gamma}_{r-}dM_r.\end{aligned}$$

For some predictable process $\bar{\gamma}$, we can use again the semi-martingale product formula to obtain:

$$\begin{aligned}\Lambda_t J_t^{ij}X_t &= \int_0^t J_r^{ij}X_r d\Lambda_r + \int_0^t\Lambda_r\langle X_r, e_i\rangle a_{ij}e_j dr\\ &\quad + \int_0^t\Lambda_r AJ_r^{ij}X_r dr + \int_0^t\Lambda_{r-}\bar{\gamma}_{r-}dM_r,\end{aligned}$$

conditioning each side on $\mathcal{Y}_t$ under the measure P_0, we see that:

$$\begin{aligned}\sigma(J_t^{ij}X_t) &= \int_0^t A\sigma(J_r^{ij}X_r)dr + \int_0^t \langle\sigma(X_r), e_i\rangle a_{ij}e_j dr \\ &\quad + \int_0^t \bar{E}(J_r^{ij}X_r d\Lambda_r|\mathcal{Y}_t).\end{aligned}$$

Recalling from (4.3.2) that

$$d\Lambda_t = \xi^{-1}\Lambda_t\langle g, X_t\rangle d\tilde{W}_t + \Lambda_{t-}[\langle\lambda, X_t\rangle - 1]d(\tilde{N}_t - t).$$

Equations (4.5.6) follows.

The derivations of $\sigma(O_t^iX_t)$, $\sigma(G_{1t}^iX_t)$ and $\sigma(G_{2t}^iX_t)$ are very similar to $\sigma(J_t^{ij}X_t)$, hence the proofs are omitted.

Theorem 4.5.2. *We have for $0 \le t$ and $1 \le i, j \le S$, $i \ne j$, the following filter equations*

$$\begin{aligned}\sigma(O_t^iX_t) &= \int_0^t \langle\sigma(X_r), e_i\rangle e_i dr + \int_0^t A\sigma(O_r^iX_r)dr \\ &\quad + \int_0^t B\sigma(O_r^iX_r)d\tilde{W}_r + \int_0^t G\sigma(O_r^iX_r)d(\tilde{N}_r - r). \qquad (4.5.7)\end{aligned}$$

$$\begin{aligned}\sigma(G_{1t}^iX_t) &= \int_0^t \langle g, e_i\rangle\langle\sigma(X_r), e_i\rangle e_i dr + \int_0^t A\sigma(G_{1r}X_r)dr \\ &\quad + \int_0^t (B\sigma(G_{1r}^iX_r) + \langle\sigma(X_r), e_i\rangle e_i)d\tilde{W}_r \\ &\quad + \int_0^t G\sigma(G_{1r}^iX_{r-})d(\tilde{N}_r - r). \qquad (4.5.8)\end{aligned}$$

$$\begin{aligned}\sigma(G_{2t}^iX_t) &= \int_0^t A\sigma(G_{2r}^iX_r)dr + \int_0^t \langle\sigma(X_r), e_i\rangle\langle\lambda, e_i\rangle e_i d\tilde{N}_r \\ &\quad + \int_0^t G\sigma(G_{2r-}^iX_{r-})d(\tilde{N}_r - r) + \int_0^t B\sigma(G_{2r}X_{r-})d(\tilde{N}_r - r). \\ &\qquad (4.5.9)\end{aligned}$$

4.6 Conclusion

In this Chapter we proposed the hidden Markov chain filtering for diffusion with jump process. This work has two main contributions to the filtering literature. Firstly, we develop the hidden Markov chain filtering methodologies to diffusion with jump model, our work generalized Elliott [30] result, and when the equation (4.2.1) only has the jump part, our work is similar to [51]. Secondly, we use the EM Algorithm in diffusion with jump model; this work is a direct extension of Dembo and Zeitouni [21], [22], Ryden [67], [68].

Chapter 5

Parameter Estimation for a Regime-Switching Mean-Reverting with Jump Model

5.1 Introduction

Brownian motion and normal distribution have been widely used to study financial derivatives and the return of assets. An extensive asset pricing theory has been developed from the contribution of Merton [59], Harrison and Pliska [42] and Harrison and Kreps [41], many authors follow Merton's (Merton [59]) work on asset price oscillations. The arrival of normal information over an infinitesimal time interval generates only marginal adjustment of the prices, which is modelled by a continuous diffusion process; whereas the arrival of abnormal information (very important news) generates a stochastic shock (jump), which is modelled as a Poisson process. Das [18], [19], Dias [25], Deng [24] and Kou [47] use mean-reverting model with jumps in the interest rate and energy market. However, recent analysis suggests that the reverting level changes have an important effect on the prices process (Hansen and Poulsen [40], Schlogl and Sommer [69]). We choose the computationally tractable model as our

basic model and augment it by letting the level to which the process reverts change. These changes are governed by a continuous time finite state Markov chain. This model is firstly put forward in the financial literature. It is then the first to use diffusion with jump hidden Markov chain filtering in this setting.

This Chapter proceeds as follows: in Section 2 the model is presented and in Section 3 the reference probability method is introduced. This is an important method for filtering and parameter estimation. In Section 4, we follow the methodology of Chapter 4 to derive a finite dimensional filter for the unobservable state of the Markov chain. This is based on observations of the mean reverting diffusion process with jumps. Various auxiliary filters will be developed that allow us to estimate the parameters of the Markov chain. In Section 5 model calibration is provided, EM algorithm is used to estimate the parameters of Markov chain. In Section 6, we will use the approximation methods which are discussed in Chapter 3, Section 3, to implement the filtering equation. Simulation results are in section 7. Finally the model is applied to Brent Oil prices. We will compare our model with the mean-reverting with jump (no regime-switching) model and the regime-switching (no jump) model in Section 8.

5.2 Regime-Switching Mean-Reverting With Jump Model

Mean reverting with jumps process is commonly used in finance. We consider a type of model, called regime-switching mean-reverting with jump, where the regime level switches according to a continuous time finite state Markov chain. This could be a

model for an interest rate or an asset price.

We consider the reference level $\bar{L} = \{\bar{L}_t; 0 \leq t \leq T\}$ a finite-state continuous time Markov chain, where $T > 0$ is a finite time horizon. Let our model for an asset price or an interest rate be described by the stochastic differential equation:

$$dL_t = \gamma(\bar{L}_t - \alpha L_t)dt + \xi dW_t + \sum_{l=1}^{n} c_l dN_t^{(l)}, \tag{5.2.1}$$

where $W = \{W_t : 0 \leq t\}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, P)$, which is independent of X_t, and each $N^{(l)} = \{N_t^{(l)} : 0 \leq t\}$, $l = 1, 2, \cdots, n$ is a Poisson process with intensities $\lambda_t^{(l)}$ and $N_t^{(j)}$ counts the number of jumps of size c_j. Note we may assume that $N_t^{(j)}$ and $N_t^{(k)}$ do not charge a common jump time if $j \neq k$, because if $N_t^{(j)}$ and $N_t^{(k)}$ have a common jump time, we can consider another $N_t^{(m)}$ with jump size $c_m = c_j + c_k$. All processes are initially defined on $(\Omega, \mathcal{F}, P)$.

In order to model the stochastic reference level $\bar{L}$ it is convenient to consider an N-state continuous time Markov chain $X = \{X_t : 0 \leq t \leq T\}$ that is identical to $\bar{L}$ after a transformation of the state space. We use the canonical representation of the finite state Markov chain introduced in [30]. That is, without loss of generality we take the state space for X to be the set $\Sigma = \{e_1, e_2, \cdots, e_S\}$, where elements e_i are column vectors with unity in the i^{th} position and zero elsewhere.

$$e_i \in \left(\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right) = \mathcal{L} \in \mathbb{R}^S$$

Then we can write,

$$\bar{L}_t = \langle \eta, X_t \rangle,$$

for an appropriate vector $\eta = (\eta_1, \cdots, \eta_S) \in \mathbb{R}^S$, where $\langle \eta, X_t \rangle$ denotes the inner product of the vector X_t and η. The essential feature of this canonical representation for the Markov chain X is that the dynamics can be written down in semi-martingale form:

$$X_t = X_0 + \int_0^t AX_r dr + M_t, \tag{5.2.2}$$

where M_t is a (vector) $\mathcal{F}_t$-martingale under P and $A \in R^{S \times S}$ is a time invariant rate matrix, whose elements are the infinitesimal intensities of X. To denote an element of the Matrix A at row i and column j, we write $\langle Ae_i, e_j \rangle$. Here $\langle \cdot, \cdot \rangle$ denotes an inner product.

We assume that the intensity is a function of X_t on a given probability space $(\Omega, \mathcal{F}, P)$.

$$\lambda^{(l)}(t) = \lambda^{(l)}(X_t) = \langle \lambda^{(l)}, X_t \rangle = \sum_{i=1}^{S} 1_{\{\omega : X_t(\omega) = e_i\}} \langle \lambda^{(l)}, e_i \rangle \in \mathbb{R}. \tag{5.2.3}$$

which denotes the number of events that occurs during the interval [0,t]. The observation process has the form:

$$dL_t = \gamma(\langle \eta, X_t \rangle - L_t)dt + \xi dW_t + \sum_{l=1}^{n} c_l dN_t^{(l)}. \tag{5.2.4}$$

We now introduce some filtrations. For the $\sigma-$subalgebras we write: $\mathcal{F}_t = \mathcal{F}_t^x \vee \mathcal{Y}_t$, where $\mathcal{F}_t^x = \sigma\{X_s : s \leq t\}$ and $\mathcal{Y}_t = \sigma\{y_s : s \leq t\}$, each contains all P-null subsets in $\mathcal{F}$. For simplicity, we make further assumptions on the independence of Poisson process and Markov processes.

We consider the situation where the process L is observed and inferences are to be made about the process $\tilde{L}$ and other parameters.

5.3 Reference Probability

Define a new probability measure P^+ such that the $\mathcal{F}_t$ restriction of the Radon-Nikodym derivative of P^+ with respect to P is:

$$\begin{aligned}\Lambda_t = \frac{dP}{dP^+}|_{\mathcal{F}_t} &= \prod_{l=1}^{n}\Big(\prod_{0<r\le t}\langle\lambda^{(l)}, X_r\rangle^{\Delta\tilde{N}_r^{(l)}}\Big)\exp\Big[\int_0^t \gamma\xi^{-1}(\langle\eta, X_r\rangle - L_r)d\tilde{W}_r \\ &\quad -\frac{1}{2}\int_0^t \gamma^2\xi^{-2}(\langle\eta, X_r\rangle - L_r)^2 dr + \sum_{l=1}^{n}\int_0^t (1-\langle\lambda^{(l)}, X_r\rangle)dr\Big].\end{aligned} \tag{5.3.1}$$

Note that: $\Lambda_t = \Lambda_t^c * \Lambda_t^d$, where

$$\Lambda_t^c = \exp\Big[\int_0^t \gamma\xi^{-1}(\langle\eta, X_r\rangle - L_r)d\tilde{W}_r - \frac{1}{2}\int_0^t \gamma^2\xi^{-2}(\langle\eta, X_r\rangle - L_r)^2 dr\Big],$$

$$\Lambda_t^d = \prod_{l=1}^{n}\Big(\prod_{0<r\le t}\langle\lambda^{(l)}, X_r\rangle^{\Delta\tilde{N}_r^{(l)}}\Big)\exp\Big[\sum_{l=1}^{n}\int_0^t (1-\langle\lambda^{(l)}, X_r\rangle)dr\Big].$$

Λ_t^c, Λ_t^d are $(\mathcal{F}_t, P^+)$ martingales with dynamics

$$\Lambda_t = 1 + \int_0^t \Lambda_r(\langle\eta, X_r\rangle - L_r)d\tilde{W}_r + \sum_{l=1}^{n}\int_0^t \Lambda_{r-}(\langle\lambda^{(l)}, X_r\rangle - 1)d(\tilde{N}_r^{(l)} - r). \tag{5.3.2}$$

Λ_t^{-1} is also a $(\mathcal{F}_t, P^+)$ martingale with dynamics

$$\Lambda_t^{-1} = 1 - \int_0^t \Lambda_r(\langle\eta, X_r\rangle - L_r)d\tilde{W}_r - \sum_{l=1}^{n}\int_0^t \Lambda_{r-}(\langle\lambda^{(l)}, X_r\rangle - 1)d(\tilde{N}_r^{(l)} - r). \tag{5.3.3}$$

Under P^+, the process $\tilde{W}_t$ is a standard Brownian motion and $\tilde{N}_t^{(l)}$ is a standard Poisson process; however, the Markov chain dynamics remains unchanged and they

are independent of each other. The state and observation process dynamics have the form:

$$P^+ \quad \begin{cases} dL_t = \xi d\tilde{W}_t + \sum_{l=1}^n c_l d\tilde{N}_t^{(l)} \\ dX_t = AX_t dt + dM_t \end{cases}$$

Under the 'real world' probability P the dynamics have the form:

$$P \quad \begin{cases} dL_t = \gamma(\langle \eta, X_r \rangle - L_t)dt + \xi dW_t + \sum_{l=1}^n c_l dN_t^{(l)} \\ dX_t = AX_t dt + dM_t \end{cases}$$

Notation: Suppose ϕ_t is an $\mathcal{F}_t$-adapted integrable process and wish to estimate $E[\phi_t|\mathcal{Y}_t]$, Using Bayes' rule (Chapter 2, Section 2),

$$\hat{\phi}_t = E[\phi_t|\mathcal{Y}_t] = \frac{\bar{E}(\Lambda_t \phi_t|\mathcal{Y}_t)}{\bar{E}(\Lambda_t|\mathcal{Y}_t)} = \frac{\sigma(\phi_t)}{\sigma(1)}.$$

5.4 Filtering Equation

In this section we derive filters for various statistics, each concerning the indirectly observed Markov process X and each computed using the observation L up to and including time t. These will be used in Section 5.5 for the E-step of EM algorithm to estimate the transition intensity matrix A and jump intensity rate. These quantities are listed below.

1. X_t, the state of the Markov chain. This process satisfies the dynamics given by equation(5.2.2), then:

$$\begin{aligned} \sigma(X_t) &= \sigma(X_0) + \int_0^t A\sigma(X_r)dr + \int_0^t B\sigma(X_r)d\tilde{W}_r \\ &\quad + \sum_{l=1}^n \int_0^t \mathrm{diag}\{\langle \lambda^{(l)}, e_i \rangle - 1\}\sigma(X_{r-})d(\tilde{N}_r^{(l)} - r). \end{aligned} \tag{5.4.1}$$

Here $\mathrm{diag}\{\langle\lambda^{(l)}, e_i\rangle - 1\}$ denotes a diagonal matrix with entries $\langle\lambda^{(l)}, e_1\rangle - 1, \cdots, \langle\lambda^{(l)}, e_S\rangle - 1$, A is the transition intensity matrix for the process X, and $B = \gamma\xi^{-1}\mathrm{diag}\{\langle\eta, e_i\rangle - L_r\}$.

2. O_t^i the amount of time spent by the process X in the state e_i up to time t:

$$O_t^i = \int_0^t \langle X_u, e_i\rangle du \in \mathbb{R}, \qquad 1 \le i \le S.$$

3. J_t^{ij} the number of transitions $e_i \to e_j$ of X where $i \neq j$, up to time t:

$$J_t^{ij} = \int_0^t \langle X_{u^-}, e_i\rangle\langle dX_u, e_j\rangle \in \mathbb{R}, \qquad 1 \le i \le S.$$

4. $\Gamma_t^{i(l)}$, the level integral for the state e_i,

$$\Gamma_t^{i(l)} = \int_0^t \langle X_u, e_i\rangle d\tilde{N}_u^{(l)} \in \mathbb{R}, \qquad 1 \le i \le S, \quad 1 \le l \le n.$$

5. G_t^i, we define below:

$$G_t^i = \int_0^t \gamma(\eta_i - L_r)\langle X_r, e_i\rangle dr + \int_0^t \xi\langle X_r, e_i\rangle d\tilde{W}_r \in \mathbb{R}, \quad 1 \le i \le S. \tag{5.4.2}$$

The measure valued quantities O_t^i, J_t^{ij}, $\Gamma_t^{i(l)}$ and G_t^i are vectors in $\mathbb{R}^n$, and satisfy the dynamics:

$$\begin{aligned}\sigma(O_t^i X_t) &= \int_0^t A\sigma(O_r^i X_r)dr + \int_0^t B\sigma(O_r^i X_r)d\tilde{W}_r \\ &\quad + \sum_{l=1}^n \int_0^t \mathrm{diag}\{\langle\lambda^{(l)}, e_i\rangle - 1\}\sigma(O_r^i X_r)d(\tilde{N}_r^{(l)} - r) \\ &\quad + \int_0^t \langle\sigma(X_r), e_i\rangle e_i dr, \end{aligned} \tag{5.4.3}$$

$$\begin{aligned}\sigma(G_t^i X_t) &= \int_0^t A\sigma(G_r^i X_r)dr + \int_0^t \gamma(\eta_i - L_r)\langle\sigma(X_r), e_i\rangle e_i dr \\ &\quad + \int_0^t (B\sigma(G_r^i X_r) + \langle\sigma(X_r), e_i\rangle e_i)d\tilde{W}_r \\ &\quad + \int_0^t \mathrm{diag}\{\langle\lambda^{(l)}, e_i\rangle - 1\}\sigma(G_r^i X_r)d(\tilde{N}^{(l)} - r). \end{aligned} \tag{5.4.4}$$

$$\begin{aligned}\sigma(\Gamma_t^{i(l)}X_t) &= \int_0^t \langle\lambda^{(l)},e_i\rangle\langle\sigma(X_r),e_i\rangle dr + \int_0^t A\sigma(\Gamma_r^{i(l)}X_r)dr \\ &\quad + \sum_{m=1}^{n}\int_0^t \mathrm{diag}\{\langle\lambda^{(m)},e_i\rangle - 1\}\sigma(\Gamma_r^{i(l)}X_r)d(N^{(m)}-r) \\ &\quad + \int_0^t \langle\lambda^{(l)},e_i\rangle\langle\sigma(X_r),e_i\rangle e_i d(N_r^{(l)}-r) + \int_0^t B\sigma(\Gamma_r^{i(l)}X_r)d\tilde{W}_r.\end{aligned} \tag{5.4.5}$$

and

$$\begin{aligned}\sigma(J_t^{ij}X_t) &= \int_0^t A\sigma(J_r^{ij}X_r)dr + \int_0^t B\sigma(J_r^{ij}X_r)d\tilde{W}_r + \int_0^t \langle\sigma(X_r),e_i\rangle a_{ji}e_j dr \\ &\quad + \sum_{l=1}^{n}\int_0^t \mathrm{diag}\{\langle\lambda^{(l)},e_i\rangle - 1\}\sigma(J_r^{ij}X_{r-})d(\tilde{N}_r^{(l)}-r).\end{aligned} \tag{5.4.6}$$

We give the proof of equation (5.4.1), the derivations of equation (5.4.3), (5.4.4), (5.4.5) and (5.4.6) are very similar to (5.4.1), hence the proofs are omitted. Similar dynamics have been established in Chapter 4 equations (4.5.6), (4.5.7), (4.5.8), (4.5.9).

Proof. In order to obtain the filtered estimate, we use generalized Itó formula:

$$\begin{aligned}\Lambda_t X_t &= X_0\Lambda_0 + \int_0^t X_{r-}d\Lambda_r + \int_0^t \Lambda_{r-}dX_r + [\Lambda,M]_t \\ &= X_0\Lambda_0 + \int_0^t X_{r-}\Lambda_r(\langle\eta,X_r\rangle - L_r)d\tilde{W}_r \\ &\quad + \sum_{l=1}^{n}\int_0^t X_{r-}\Lambda_r(\langle\lambda^{(l)},X_r\rangle - 1)d(\tilde{N}_r^{(l)}-r) \\ &\quad + \int_0^t \Lambda_{r-}AX_r dr + \int_0^t \Lambda_{r-}dM_r + [\Lambda,M]_t,\end{aligned}$$

for $t \geq 0$, conditioning each side of the equation above on $\mathcal{Y}_t$, we obtain that:

$$\begin{aligned}\sigma(X_t) &= \sigma(X_0) + \int_0^t A\sigma(X_r)dr + \int_0^t \bar{E}(X_r\Lambda_r(\langle\eta,X_r\rangle - L_r)|\mathcal{Y}_r)d\tilde{W}_r \\ &\quad + \int_0^t \bar{E}(X_{r-}\Lambda_{r-}(\langle\lambda^{(l)},X_r\rangle - 1)|\mathcal{Y}_r)d(\tilde{N}_r^{(l)}-r)\end{aligned}$$

Note that: $\sum_{i=1}^{S}\langle X_r, e_i\rangle = 1$, we have:

$$\bar{E}(X_r\Lambda_r(\langle \eta, X_r\rangle - L_r)|\mathcal{Y}_r) = \sum_{i=1}^{S} e_i(\langle \eta, e_i\rangle - L_r)\langle \sigma(X_r), e_i\rangle.$$

The same argument, we have:

$$\bar{E}(X_r\Lambda_{r^-}(\langle \lambda^{(l)}, X_r\rangle - 1)|\mathcal{Y}_t) = \sum_{i=1}^{S} e_i(\langle \lambda^{(l)}, e_i\rangle - 1)\langle \sigma(X_r), e_i\rangle.$$

(5.4.1) follows.

□

5.5 Model Calibration

Most financial models always depend on one or several parameters. The problem of calibration is to provide the value of these parameters with two constraints of matching the market price of liquid instruments and of fitting certain statistics.

The filtering methodology that we describe is based on a change of measure technique which by its nature does not provide direct estimates of diffusion coefficients. Various forms and methods of estimating the volatility or diffusion coefficient can be found in the literature; see for example (Rogers and Satchell [66]). We will suppose ξ and c_l are constants determined by one of these techniques. Using filtering results and hidden Markov models we wish to describe how $\bar{L}_t$ and its stochastic behavior can be estimated. We also wish to estimate the transition intensity matrix A and the vector $\lambda^{(l)}$ of jump intensity rate. Suppose we have observations $\{L_1, \cdots, L_T\}$ available, where T is a fixed positive integer. Often, the likelihood function is difficult to compute directly, in many cases, the EM algorithm is used.

Suppose the model is first estimated using a set of parameters $\theta' = \{a'_{ij}, \lambda_i^{'(l)}; 1 \leq i,j \leq N, 1 \leq l \leq n\}$ and we wish to determine a new set $\theta = \{a_{ij}, \lambda_i^{(l)}; 1 \leq i,j \leq N, 1 \leq l \leq n\}$ which maximizes the log-likelihood function defined below.

Let J_t^{ij} denote the number of jumps that the process X makes state e_i to e_j in the interval [0,t]. We define:

$$\begin{aligned}
\frac{dP_\theta}{dP_{\theta'}} &= \exp\Big[-\int_0^t \gamma\xi^{-1}(\langle\eta, X_r\rangle - \langle\eta', X_r\rangle)d\tilde{W}_r \\
&\quad -\frac{1}{2}\int_0^t (\gamma^2\xi^{-2}(\langle\eta, X_r\rangle - L_r)^2 - \gamma^2\xi^{-2}(\langle\eta', X_r\rangle - L_r)^2)dr \\
&\quad +\sum_{l=1}^{n}\Big(\int_0^t (\log\frac{\langle\lambda^{(l)}, X_r\rangle}{\langle\lambda'^{(l)}, X_r\rangle})d\tilde{N}_r^{(l)} - \int_0^t \langle\lambda^{(l)} - \lambda'^{(l)}, X_r\rangle dr\Big)\Big] \\
&\quad \times \prod_{i,j=1, i\neq j}^{S} \exp\Big[\int_0^t \log(\frac{a_{ji}}{a'_{ji}})dJ_r^{ij} + \Big(\int_0^t (a'_{ji} - a_{ji})\langle X_r, e_i\rangle dr\Big)\Big].
\end{aligned} \tag{5.5.1}$$

We have:

$$\begin{aligned}
\log(\frac{dP_\theta}{dP_{\theta'}}) &= -\int_0^t \xi^{-1}\langle\eta - \eta', X_r\rangle d\tilde{W}_r \\
&\quad -\frac{1}{2}\int_0^t (\gamma^2\xi^{-2}(\langle\eta, X_r\rangle - L_r)^2 - \gamma^2\xi^{-2}(\langle\eta', X_r\rangle - L_r)^2)dr \\
&\quad +\sum_{l=1}^{n}\int_0^t (\log\langle\lambda^{(l)}, X_r\rangle - \log\langle\lambda'^{(l)}, X_r\rangle)d\tilde{N}_r^{(l)} - \int_0^t \langle\lambda^{(l)} - \lambda'^{(l)}, X_r\rangle dr \\
&\quad +\sum_{i,j=1, i\neq j}^{S} (J_t^{ij}(\log a_{ji} - \log a'_{ji}) + \int_0^t (a'_{ji} - a_{ji})\langle X_r, e_i\rangle dr),
\end{aligned} \tag{5.5.2}$$

The parameter estimation can be extended to also estimate the speed of adjustment γ. The procedure used to estimate the parameters a_{ij}, $1 \leq i,j \leq N$, remains

unchanged. Including γ as a parameter that we wish to estimate implies that equation (5.5.1) takes a different form. The Radon-Nikodym derivative of the probability measure induced by the update parameter value $\theta = \{\gamma, a_{ij}, \lambda_i^{(l)}; 1 \leq i, j \leq S, 1 \leq l \leq n\}$ with respect to the probability measure induced by the old parameter values $\theta' = \{\gamma', a'_{ij}, \lambda_i^{'(l)}; 1 \leq i, j \leq S, 1 \leq l \leq n\}$ is then

$$\begin{aligned}
\frac{dP_\theta}{dP_{\theta'}} &= \exp\big(-\int_0^t \gamma\xi^{-1}(\langle\eta, X_r\rangle - L_r) - \gamma'\xi^{-1}(\langle\eta', X_r\rangle - L_r))d\tilde{W}_r \\
&\quad -\frac{1}{2}\int_0^t (\gamma^2\xi^{-2}(\langle\eta, X_r\rangle - L_r)^2 - \gamma'^2\xi^{-2}(\langle\eta', X_r\rangle - L_r)^2)dr \\
&\quad +\sum_{l=1}^{n}\Big(\int_0^t (\log\langle\lambda^{(l)}, X_r\rangle - \log\langle\lambda^{'(l)}, X_r\rangle)d\tilde{N}_r^{(l)} - \int_0^t \langle\lambda^{(l)} - \lambda^{'(l)}, X_r\rangle dr\Big) \\
&\quad + \sum_{i,j=1, i\neq j}^{S} (J_t^{ij}(\log a_{ji} - \log a'_{ji}) + \int_0^t (a'_{ji} - a_{ji})\langle X_r, e_i\rangle dr). \qquad (5.5.3)
\end{aligned}$$

Here, we take the conditional expectation on $\mathcal{Y}_t$ of both sides (5.5.2) and obtain:

$$\begin{aligned}
Q(\theta, \theta') &= \sum_{i=1}^{S}\gamma\xi^{-1}\eta_i E\Big(\int_0^t \langle X_r, e_i\rangle d\tilde{W}_r|\mathcal{Y}_t\Big) \\
&\quad -\sum_{i=1}^{S}\frac{1}{2}\gamma^2\xi^{-2}\eta_i^2 E\Big(\int_0^t (\langle X_r, e_i\rangle - 2L_r\langle X_r, e_i\rangle)dr|\mathcal{Y}_t\Big) \\
&\quad +\sum_{l=1}^{n}\sum_{i=1}^{S}\Big(\log\lambda_i^{(l)} E\Big(\int_0^t \log\langle X_r, e_i\rangle d\tilde{N}_r^{(l)}|\mathcal{Y}_t\Big) - \lambda_i^{(l)} E\Big(\int_0^t \langle X_r, e_i\rangle dr|\mathcal{Y}_t\Big)\Big) \\
&\quad + \sum_{i,j=1, i\neq j}^{S}\Big(\log a_{ji} E(J_t^{ij}|\mathcal{Y}_t) - a_{ji} E\Big(\int_0^t \langle X_r, e_i\rangle dr|\mathcal{Y}_t\Big) + R(\theta')\Big), \\
&\qquad (5.5.4)
\end{aligned}$$

where the term $R(\theta')$ does not involve the parameter θ, so the E-step consists of computing $\hat{J}_t^{ij}$, $\hat{O}_t^i$ and $\hat{G}_t^i$.

For the M-step, set the partial derivatives with respect to a_{ji}, η_i and $\lambda_i^{(l)}$ to zero,

this yields the estimates $\hat{\theta}$ as

$$\hat{a}_{ji} = \frac{E(J_t^{ij}|\mathcal{Y}_t)}{E(\int_0^t \langle X_r, e_i\rangle dr|\mathcal{Y}_t)}, \tag{5.5.5}$$

$$\hat{\lambda}_i^{(l)} = \frac{E(\int_0^t \langle X_r, e_i\rangle d\tilde{N}_r^{(l)}|\mathcal{Y}_t)}{E(\int_0^t \langle X_r, e_i\rangle dr|\mathcal{Y}_t)}, \tag{5.5.6}$$

$$\hat{\eta}_i = \frac{\gamma^{-1}\xi E(\int_0^t \langle X_r, e_i\rangle d\tilde{W}_r|\mathcal{Y}_t) + E(\int_0^t L_r\langle X_r, e_i\rangle dr|\mathcal{Y}_t)}{E(\int_0^t \langle X_r, e_i\rangle dr|\mathcal{Y}_t)}. \tag{5.5.7}$$

One of the basic properties of the transition intensity matrix A is that for each $i = 1, \cdots, M$, $\sum_{j=1}^{M} a_{ji} = 0$, so that there is no need to estimate the elements $a_{ii}, i = 1, \cdots, M$, if the others are already estimated.

5.6 Implementation of Filter

The numerical algorithm to solve the stochastic differential equation about diffusion with jump process has been discussed in Chapter 3 Section 3. As noted above, the unnormalized filtering equations are preferable because they are linear and driven by the observation L.

To obtain an approximation for equation (5.4.1), we discretize the dynamics for the process $\sigma(X_t)$ over a regular partition. We consider the interval [0,t]. Let $q_t = \sigma(X_t)$, then equation (5.4.1) is:

$$\begin{aligned} q_t &= q_0 + \int_0^t Aq_r dr + \int_0^t Bq_r d\tilde{W}_r \\ &\quad + \sum_{l=1}^{n} \int_0^t \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1\} q_{r-} d(\tilde{N}_r^{(l)} - r). \end{aligned} \tag{5.6.1}$$

Then,

$$P(X_t = e_i|\mathcal{Y}_t) = \frac{\langle q_t, e_i\rangle}{\sum_{m=1}^{S}\langle q_t, e_m\rangle}.$$

Suppose $h = \frac{t}{n}$, for $0 \leq k < n$.

The first approximation (Chapter 3 Section 3.3)gives:

$$\begin{aligned} q_{(k+1)h} &= q_{kh} + Aq_{kh}h + Bq_{kh}(\tilde{W}_{(k+1)h} - \tilde{W}_{kh}) \\ &\quad + \sum_{l=1}^{n} \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1\} q_{kh}(\tilde{N}^{(l)}_{(k+1)h} - \tilde{N}^{(l)}_{kh} - h). \end{aligned} \tag{5.6.2}$$

The second approximation gives:

$$\begin{aligned} q_{(k+1)h} &= q_{kh} + Aq_{kh}h + Bq_{kh}(\tilde{W}_{(k+1)h} - \tilde{W}_{kh}) \\ &\quad + \sum_{l=1}^{n} \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1\} q_{kh}(\tilde{N}^{(l)}_{(k+1)h} - \tilde{N}^{(l)}_{kh} - h) \\ &\quad + \frac{1}{2}B^2 q_{kh}((\tilde{W}_{(k+1)h} - \tilde{W}_{kh})^2 - h). \end{aligned} \tag{5.6.3}$$

Write $\tau^i_t = \sigma(O^i_t X_t)$, then (5.4.3),

The first approximation gives:

$$\begin{aligned} \tau^i_{(k+1)h} &= \tau^i_{kh} + A\tau^i_{kh}h + B\tau^i_{kh}(\tilde{W}_{(k+1)h} - \tilde{W}_{kh}) \\ &\quad + \sum_{l=1}^{n} \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1)\} \tau^i_{kh}(\tilde{N}^{(l)}_{(k+1)h} - \tilde{N}^{(l)}_{kh} - h) + \langle q_{kh}, e_i\rangle e_i h. \end{aligned} \tag{5.6.4}$$

The second approximation gives:

$$\begin{aligned} \tau^i_{(k+1)h} &= \tau^i_{kh} + A\tau^i_{kh}h + B\tau^{ij}_{kh}(\tilde{W}_{(k+1)h} - \tilde{W}_{kh}) \\ &\quad + \sum_{l=1}^{n} \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1\} \tau^i_{kh}(\tilde{N}^{(l)}_{(k+1)h} - \tilde{N}^{(l)}_{kh} - h) + \langle q_{kh}, e_i\rangle e_i h \\ &\quad + \frac{1}{2}B^2 \tau^i_{kh}((\tilde{W}_{(k+1)h} - \tilde{W}_{kh})^2 - h). \end{aligned} \tag{5.6.5}$$

Write $n_t^{ij} = \sigma(J_t^{ij} X_t)$, then (5.4.6),

The first approximation gives:

$$\begin{aligned} n_{(k+1)h}^{ij} &= n_{kh}^{ij} + An_{kh}^{ij}h + Bn_{kh}^{ij}(\tilde{W}_{(k+1)h} - \tilde{W}_{kh}) + \langle q_{kh}, e_i\rangle a_{ji}e_j h \\ &\quad + \sum_{l=1}^{n} \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1)\} n_{kh}^{ij}(\tilde{N}_{(k+1)h}^{(l)} - \tilde{N}_{kh}^{(l)} - h). \end{aligned} \tag{5.6.6}$$

The second approximation gives:

$$\begin{aligned} n_{(k+1)h}^{ij} &= n_{kh}^{ij} + An_{kh}^{ij}h + Bn_{kh}^{ij}(\tilde{W}_{(k+1)h} - \tilde{W}_{kh}) + \langle q_{kh}, e_i\rangle a_{ji}e_j h \\ &\quad + \sum_{l=1}^{n} \mathrm{diag}\{\langle \lambda^{(l)}, e_i\rangle - 1)\} n_{kh}^{ij}(\tilde{N}_{(k+1)h}^{(l)} - \tilde{N}_{kh}^{(l)} - h) \\ &\quad + \frac{1}{2}B^2 n_{kh}^{ij}((\tilde{W}_{(k+1)h} - \tilde{W}_{kh})^2 - h). \end{aligned} \tag{5.6.7}$$

New estimates for the parameter a_{ji} is $\hat{a}_{m+1,ji} = \sigma(J_t^{ij})/\sigma(O_t^i) = \langle n_{m+1}^{ij}, \mathbf{1}\rangle / \langle \tau_{m+1}^i, \mathbf{1}\rangle$, here $\mathbf{1} = (1, 1, \cdots, 1)$.

5.7 A Simulation Example

For our simulation example, we consider two Poisson jumps in our model. The filters are calculated using a standard numerical method which is discussed in section 6. To simulate asset price sample paths, we use the approximation discussed in Chapter 3. The model (5.2.4) can be discretized as:

$$L_{k+1} \simeq L_k + \gamma(\langle \eta, X_k\rangle - L_k)\Delta t_k + \xi \Delta W_k + \sum_{l=1}^{2} c_l \Delta N_k^{(l)}. \tag{5.7.1}$$

Here W_k is a Gaussian process with $\Delta W \sim N(0,1)$, $N_k^{(l)}$ is Poisson process with $\Delta N^{(l)} \sim P(\lambda^{(l)})$. In the simulation the process X_k is a two state continuous time

Markov chain $\langle \eta, e_1 \rangle = -1$ and $\langle \eta, e_1 \rangle = 1$ with initial probability vector $p(0) = (\frac{1}{2}, \frac{1}{2})$. The transition intensity matrix used in our simulations is set at

$$A = \begin{bmatrix} -0.8 & 0.5 \\ 0.8 & -0.5 \end{bmatrix},$$

the other parameter values are taken to be $\xi = 1.2$, $\gamma = 8.0$, $\langle \lambda^{(1)}, e_1 \rangle = 0.3$, $\langle \lambda^{(2)}, e_2 \rangle = 1.9$, $c_1 = 0.125$, $c_2 = -0.125$. The process L is simulated at the observed time interval [0, 30] using a total of 30,000 discretisation points.

We give some explanations about simulating asset price sample paths below. In Chapter 3, section 1, we have introduced the simulation of the continuous time finite state Markov chain, normal random generator and two-point random variable generator, therefore $\langle \eta, X_k \rangle$ and ΔW_k in the equation (5.7.1) can easily be simulated. For $\Delta N_k^{(l)}$ in the equation (5.7.1), we have:

$$\Delta N_k^{(l)} = \begin{cases} 1, & \text{with probability} \quad \lambda^{(l)} \Delta t_k \\ 0, & \text{with probability} \quad 1 - \lambda^{(l)} \Delta t_k \end{cases}$$

In this setting, we can use two-point random variable to simulate $\Delta N_k^{(l)}$, $l = 1, 2$. (In Matlab, we can use Matlab function called randn to obtain normal random generator with mean 0 and variance 1).

The main steps in our algorithms are as follows:

Step 1 For $i, j = 1, 2$, choose the value of $a_{0,ji}$. For example, $a_{0,11} = -5.0$, $a_{0,21} = 5.0$, $a_{0,12} = 5.0$, $a_{0,22} = -5.0$.

Step 2 Using the discrete time forwards equations (5.6.2), (5.6.4), (5.6.6) and (5.5.5) to compute new estimated value $a_{m,ji}$. From (5.5.5) and Bayes' formula (Theorem 2.2.8) we have:

$$\hat{a}_{m+1,ji} = \frac{E_m(J_t^{ij}\Lambda_t|\mathcal{Y}_t)}{E_m(\int_0^t\langle X_r, e_i\rangle\Lambda_t dr|\mathcal{Y}_t)}, \tag{5.7.2}$$

namely,

$$\hat{a}_{m+1,ji} = \frac{\sigma(J_t^{ij})}{\sigma(O_t^i)} = \frac{\langle\sigma(J_t^{ij}X_t), \mathbf{1}\rangle}{\langle\sigma(O_t^iX_t), \mathbf{1}\rangle}, \tag{5.7.3}$$

here $\mathbf{1} = (1, 1, \cdots, 1)$. $\sigma(J_t^{ij}X_t)$ and $\sigma(O_t^iX_t)$ can be calculated from (5.6.6) and (5.6.4).

Step 3 When $|\hat{a}_{m+1,ji} - \hat{a}_{m,ji}| \leq 10^{-4}$, or $N = 3 \times 10^4$ loops the programme stops.

Table (5.1) shows that the EM algorithm converges to the true value, even from far away initial value. Table (5.2) demonstrates when we take different sample paths the biases of estimated parameters are very small and insignificant. It should be noted that the likelihood function possesses several local maxima, and the estimated value of the EM algorithm converges to one of these local maxima. If there are many unknown parameters in the model, the optimization method is needed in order to obtain the satisfied results.

Figure 5.1 and Figure 5.2 show the evolution of the estimates of the parameters. It is seen that the EM algorithm converges. For short observation time, the ML (maximum likelihood) estimate is not very close to the true parameter. However, collecting more data, the ML estimate becomes close to the true value. The simulation

is conducted over a long time interval to ensure that each state in the Markov chain is visited a sufficiently large number of times. Accurate estimation of the elements of the matrix depends on the observation of numerous jumps in the Markov chain.

In our numerical study, we feel that EM algorithm is very useful estimating the parameter of the model. It is convergent, stationary and less dependent on initial value especially in two state continuous time Markov chain.

Initial value	True Value	Estimation	Standard Deviation
$a_{11} = -1.5$	$a_{11} = -0.8$	$a_{11} = -0.8061$	0.3008
$a_{22} = -1.5$	$a_{22} = -0.5$	$a_{22} = -0.5028$	0.4029
$a_{11} = -5.0$	$a_{11} = -0.8$	$a_{11} = -0.7496$	0.2609
$a_{22} = -5.0$	$a_{22} = -0.5$	$a_{22} = -0.4829$	0.3953
$a_{11} = -10.0$	$a_{11} = -0.8$	$a_{11} = -0.8386$	0.3522
$a_{22} = -10.0$	$a_{22} = -0.5$	$a_{22} = -0.5329$	0.4838
$a_{11} = -15.0$	$a_{11} = -0.8$	$a_{11} = -1.5737$	0.8735
$a_{22} = -15.0$	$a_{22} = -0.5$	$a_{22} = -1.2475$	1.0169

Table 5.1: Different initial value

Sample path number	True Value	Estimation	Standard Deviation
M=20	a_{11}=-0.8	a_{11}=-0.9176	0.2614
	a_{22}=-0.5	a_{22}=-0.6402	0.4357
M=50	a_{11}=-0.8	a_{11}=-0.7496	0.2609
	a_{22}=-0.5	a_{22}=-0.4829	0.3953
M=100	a_{11}=-0.8	a_{11}=-0.8262	0.3156
	a_{22}=-0.5	a_{22} =-0.5637	0.4313
M=150	a_{11}=-0.8	a_{11}=-0.7946	0.3008
	a_{11}=-0.5	a_{22}=-0.5207	0.4007
M=200	a_{11}=-0.8	a_{11}=-0.8061	0.2224
	a_{11}=-0.5	a_{22}=-0.5128	0.3029

Table 5.2: Different simulation size M

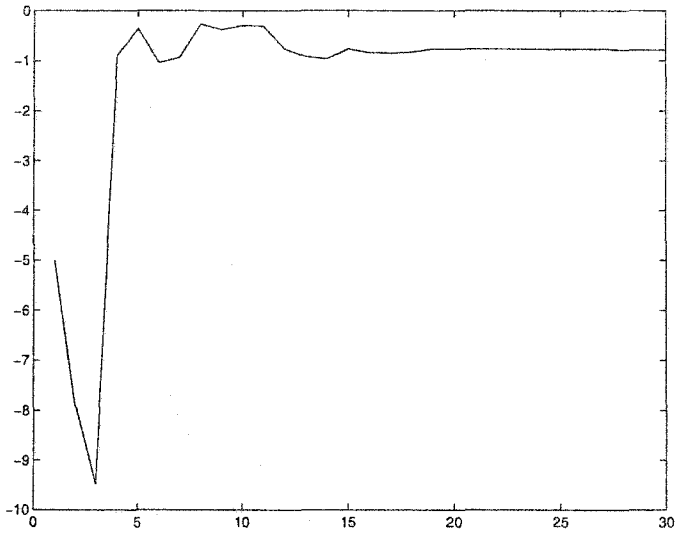

Figure 5.1: Evolution of the estimates of the parameter a_{11}

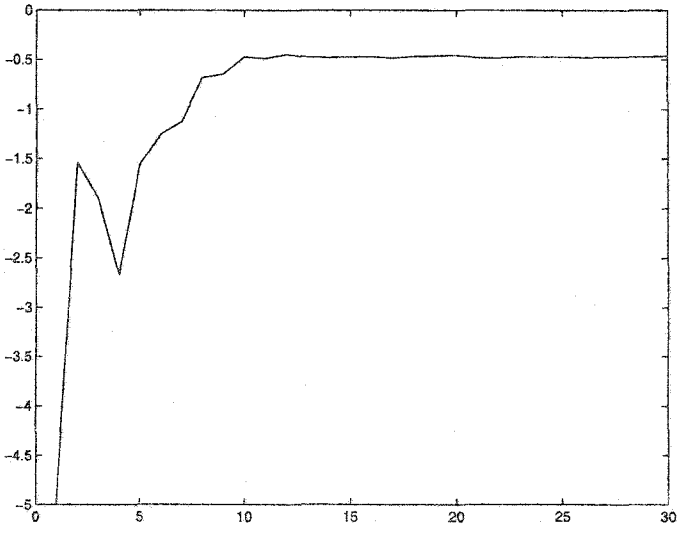

Figure 5.2: Evolution of the estimates of the parameter a_{22}

5.8 Application

To support the plausibility of using the process L in financial model as defined in equation (5.2.3), with a hidden Markov chain as a model for the reference level, we compare the Brent Oil Prices (monthly) with some typical sample paths for the process L. We feel that this model is applicable in situations where some natural reference level exists for oil prices.

Figure 5.3 is a plot on Brent oil prices (monthly). The observations are monthly observations for the period between the January 1970 and August 2000. At least four large jump-ups and three jump-downs for oil prices can be identified in these events. (Jumps in 1973/1974 Iom Kipur war and Arabian oil embargo, in 1979/1980 Iran revolution and Iran-Iraq war, in 1986 Saudi Arabia price war, in 1990 Kuwait invasion by Iraq, in 1991 the Iraq defeat, in 1997 Asian crisis, and in 1999 OPEC and allies supply shock). Figure 5.4 is the Normal Quantile Plot. Figure 5.3 can be compared with Figure 5.5 which is a simulated realization of the process L. We also plot the picture if no jumps or Regime-switching occurred in the process L, see Figure 5.6 and Figure 5.7.

We can in principle apply the techniques used in Section 5.7 to the Brent oil price dataset but it has to be made clear that, from a data series where there are only seven jumps and more than three states in the Markov chain between the period it is difficult to model calibration, unless they can be reduced to a small number of parameters that are relevant for each state of the system. This could be achieved by specifying the matrix A and jump sizes in terms of a small number of parameters.

The above techniques could be directly applied with modifications. Since the model is characterized by an independent finite state Markov chain in combination with a conditionally Gaussian and Poisson jumps observation process. It is straightforward to compute financial derivatives and value at risk estimates following standard methods.

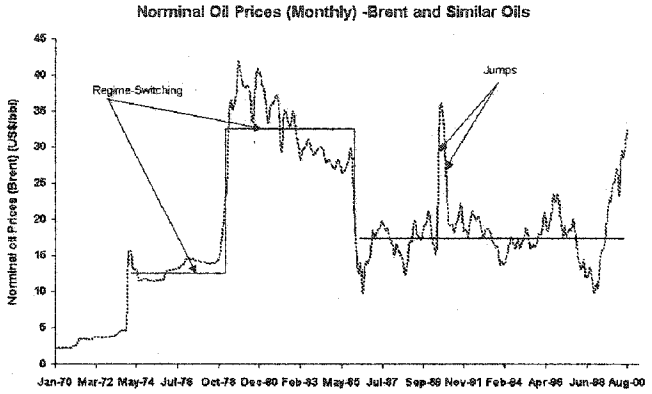

Figure 5.3: Brent Oil Prices monthly

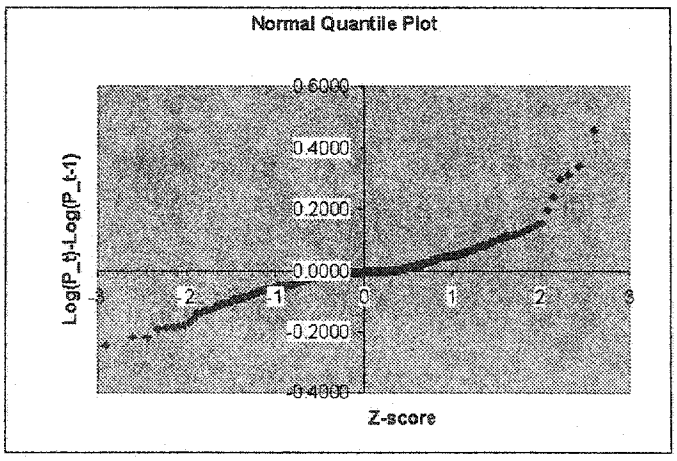

Figure 5.4: Normal Quantile Plot

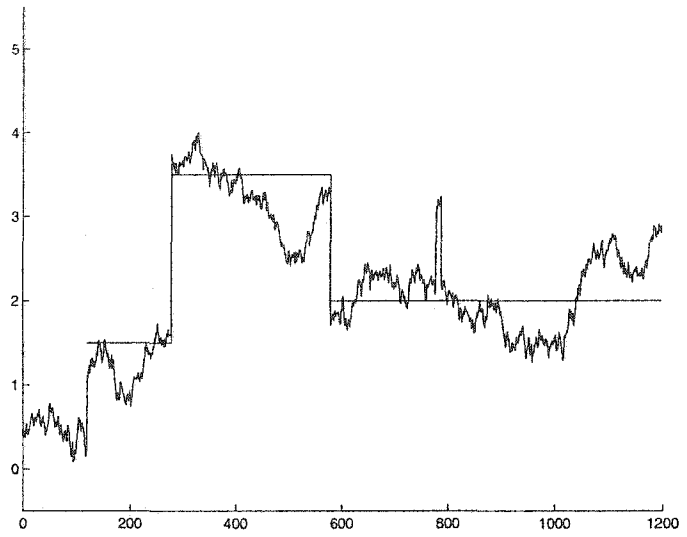

Figure 5.5: Regime-Switching Mean-Reverting with Jump Model

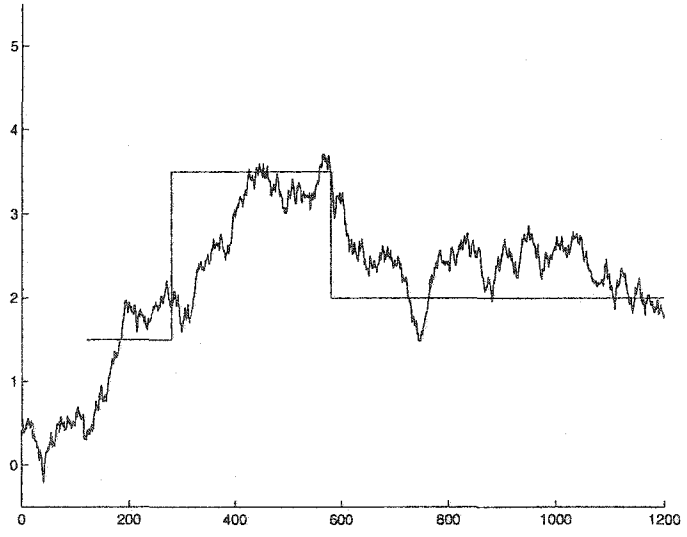

Figure 5.6: Regime-Switching Mean-Reverting Model

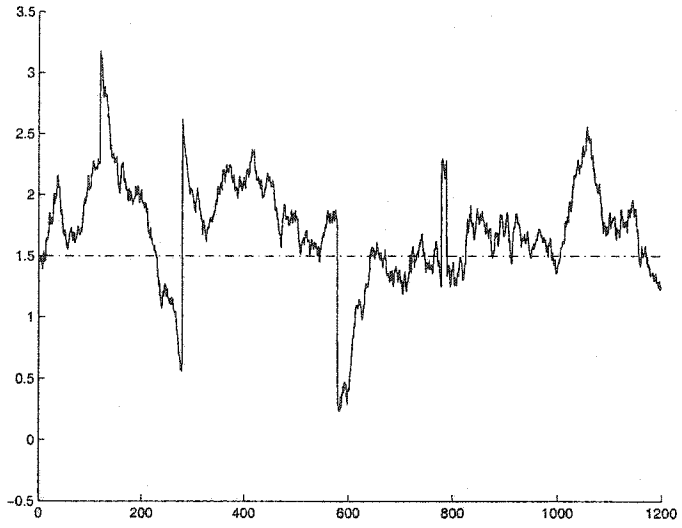

Figure 5.7: Mean-Reverting with Jump Model

Bibliography

[1] M. Abramowitz and N. C. Stegun. *Handbook of Mathematical Functions.* New York: Dower Publications Inc., 1970.

[2] S. H. Babbs and K. B. Nowman. Kalman filtering of generalized vasicek term structure models. *Journal of Financial and Quantitative Analysis*, 34(1):115–130, March 1999.

[3] D. J. Ballantyne, H. Y. Chan, and M. A. Kouritzin. A novel branching particle method for tracking. *In Signal and Data processing of small Targets, Proceedings of SPIE, 4048*, pages 277–287, 2000. Ed. O. E. Drumond.

[4] L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of American Mathematical Society*, 73:360–363, 1967.

[5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41:164–171, 1970.

[6] J. D. Beasley and S. G. Springer. Approximation to inverse normal. *Journal of Applied Statistics*, 26:118–121, 1977.

[7] V. Benes. Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics*, 5:65–92, 1981.

[8] P. Brémaud. Point processes and queues. *Springer-Verlag*, 1981.

[9] P. Brandimarte. *Numerical Methods in Finance A MATLAB-Based Introduction.* John Wiley and Sons, Inc., 2002.

[10] R. S. Bucy. Recent results in linear and non-linear filtering. *Stochastic Problems in Control, The American Society of Mechanical Engineers*, pages 87–105, 1968.

[11] M. Chaleyat-Maurel and D. Michel. Des resultats de non-existence de filtre de dimension finie. *Stochastics*, 13:83–102, 1984.

[12] Ren-Raw Chen and L. Scott. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income*, 3:14–31, 1993.

[13] Ren-Raw Chen and L. Scott. Multi-factor cox-ingersoll-ross models of the term structure: Estimates and tests from a Kalman filter model. Rutgers University, working paper, 1997.

[14] M. Chesney and R. J. Elliott. Estimating the instantaneous volatility and convariance of risky assets. *Applied Stochastic Models and Data Analysis*, 11:51–58, 1995.

[15] S. Claessens and G. G. Pennacchi. Estimating the likelihood of mexican default from market prices of brady bonds. *Journal of Financial and Quantitative Aanlysis*, 31(1):109–126, 1996.

[16] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–406, 1985.

[17] R. E. Cumby and M. D. D. Evans. The term structure of credit risk: Estimations and specification tests. New York University, working paper, 1995.

[18] S. R. Das. Poisson-Gaussian processes and the bond markets. Harvard University, working paper, 1998.

[19] S. R. Das. The surprise elemnet: Jumps in interest rate diffusions. Harvard University, working paper, 1999.

[20] M. Davis. New approach to filtering for non-linear system. *Institute of Electrical and Electronics Engineers. Proceedings*, 128(5):166–172, 1981. Part D.

[21] A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Process and their Application*, 23:91–113, 1986.

[22] A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous stochastic processes via the EM algorithm. *Stochastic Process and their Application*, 31:167–169, 1989.

[23] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1–22, 1977.

[24] S. Deng. Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes. University of California at Berkeley, working paper, 1998.

[25] M. A. G. Dias. Petroleum concessions with extendible options: Investment timing and value using mean reversion and jump processes for oil prices. working paper, 1999.

[26] Jin-Chuan Duan and Jean-Guy Simonato. Estimating and testing exponential-affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting*, 13(2):111–135, 1999.

[27] T. E. Duncan. *Doctoral Dissertation.* PhD thesis, Department of Electrical Engineering, Stanford University, 1967.

[28] R. Elliott, M. Kouritzin, T. Kurtz, and H. Long (Organizer). The conference on filtering theory and applications, July 2002. University of Alberta, Canada, $http://www.math.ualberta.ca/Kouritzin_M.html$.

[29] R. J. Elliott. *Stochastic Calculus and Application.* Springer-Verlag, 1982.

[30] R. J. Elliott. New finite-dimensional filters and smoothers for noisily observed Markov chains. *IEEE Transactions on Information Theory*, 39(1):265–271, 1993.

[31] R. J. Elliott. Exact adaptive filters for Markov chains observed in gaussian noise. *Automatica*, 30(9), 1994.

[32] R. J. Elliott, P. Fischer, and E. Platen. Filtering and parameter estimation for a mean reverting interest rate model. *Canadian Applied Mathematics Quarterly*, 7(4):1–20, Winter 1999.

[33] R. J. Elliott, J. Hoek, and J. Valencia. Non-linear filter estimation of volatility. University of Alberta, working paper, 1999.

[34] R. J. Elliott, W. C. Hunter, and B. M. Jamieson. Drift and volitility estimation in discrete time. *Journal of Economics and Control*, 22:209–218, 1998.

[35] R. J. Elliott and V. Krishnamurthy. New finite-dimensional filters for parameter estimation of discrete-time linear gaussian models. *IEEE Transactions on Automatic Control*, 44(5):938–951, May 1999.

[36] R. J. Elliott and R. W. Rishel. Estimating the implicit interest rate of a risk asset. *Stochastic Processes and Their Applications*, 49:199–206, 1994.

[37] R. Frey and W. J. Runggaldier. Risk-mimnimizing hedging strategies under restricted information: The case of stochastic volatility models observable only

at discrete random times. *Mathematical Methods of Operations Research*, 50:339–350, 1999. Springer-Verlag.

[38] S. Friihwirth-Schnatter. Applied state space modelling of non-gaussian time series using integration-based Kalman filtering. *Statistics and Computing*, 4:259–269, 1994.

[39] M. Fujisaki, G. Kallianpur, and H. Kunita. Stochastic differential equation for the nonlinear filtering problem. *Osaka Journal Mathematics*, pages 19–40, 1972.

[40] A. T. Hansen and R. Poulsen. A simple regime switching term structure model. *Finance and Stochastics*, 4:409–429, 2000.

[41] J. M. Harrison and D. M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

[42] J. M. Harrison and S. R. Pliska. Martingles and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11:215–260, 1981.

[43] A. Harvey, E. Ruig, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994.

[44] J. Hull and A. White. The pricing of options on assets with stochastic volativity. *Journal of Finance*, 42(2):281–300, 1987.

[45] G. Kallianpur. Some recent developments in nonlinear filtering theory. *Acta Applicandae Mathematica*, 1(4), 1983.

[46] S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361–393, 1998.

[47] S. G. Kou. A jump diffusion model for option pricing with three properties: Leptokurtic feature, volatility smile, and analytical tractability. Columbia University, working paper, 1999.

[48] M. A. Kouritzin. On exact filters for continuous signals with discrete observations. *IEEE Transactions on automatic control*, 43(5):709–715, May 1998.

[49] M. A. Kouritzin. Exact infinite dimensional filters and explicit solutions. *In Stochastic Models*, pages 265–282, 2000. Eds. L. G. Gorostiza and B.G. Ivanoff.

[50] M. A. Kouritzin, H. Long, and W. Sun. Markov chain approximations to filtering equations for reflecting diffusion process. Department of Mathematics and Statistics Sciences, University of Alberta, Canada, 2002.

[51] V. Krishnamurthy and R. J. Elliott. Filters for estimating Markov modulated Poisson processes and image-based tracking. *Automatica*, 33(5):821–833, 1997.

[52] H. J. Kushner. Dynamics equation for optimal nonlinear filtering. *Journal of Differential Equations*, 3:179–190, 1967.

[53] H. J. Kushner. A robust discrete state approximation to the optimal nonlinear filter for a diffusion. *Stochastics*, 3:75–83, 1979.

[54] J. Lund. Non-linear Kalman filtering techniques for term-structure models. The Aarhus School of Business, working paper, 1997.

[55] M. Manoliu and S. Tompaidis. Energy futures prices: Term structure models with Kalman filter estimation. University of Texas at Austin, working paper, 2000.

[56] G. B. Di Masi and W. J. Runggaldier. Continuous-time approximations for the nonlinear filtering problem. *Applied Mathematics and Optimization*, 7:233–245, 1981.

[57] G. B. Di Masi and W.J. Runggaldier. Continuous-time approximations for the nonlinear filtering problem. *Journal of Applied Mathematics and Optimization*, 7:233–245, 1981.

[58] A. Melino and S. M. Turnbull. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45:239–265, 1990.

[59] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

[60] D. Del Moral, M. A. Kouritzin, and L. Miclo. On a class of discrete generation interacting paricle systems. *Electronic Journal of Probability*, 6:1–26, 2001. Paper no. 16.

[61] R. E. Mortensen. *Doctoral Dissertation.* PhD thesis, Department of Electronical Engineering, University of California, Berkeley, CA, 1966.

[62] B. Oksendal. *Stochastic Differencial Equations An Introduction with Applications.* Springer, 1998.

[63] E. Pardoux. *Smoothing of Diffusion Process Conditionned at Final Time*, volume 43, pages 187–196. Springer, 1982.

[64] G. G. Pennacchi. Identifying the dynamics of real interest rates and inflation: Evidence using survey data. *Review of Financial Studies*, 4(1):53–86, 1991.

[65] P. Protter. *Stochastic Intergration and Differential Equations A New Approach.* Springer-Verlag, 1990.

[66] L. C. G. Rogers and S. E. Satchell. Estimating variance form high, low and closing prices. *Annals of Applied Probability*, 1:504–512, 1991.

[67] T. Ryden. Parameter estimation for Markov modulated Poisson process. *Stochastic Models*, 10:795–829, 1994.

[68] T. Ryden. An EM algorithm for estimation in Markov modulated Poisson process. *Computational Statistics and Data Analysis*, 21:431–447, 1996.

[69] E. Schlogl and D. Sommer. Factor models and the shape of the term structure. University of Bonn, working paper, 1997.

[70] E. S. Schwartz. The stochastic bechavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*, 52(3):923–973, July 1997.

[71] L. O. Scott. Option pricing when variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis*, 22(4):419–439, December 1987.

[72] I. M. Sobol. On the distribution of points in a cube and the approximation evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7:86–112, 1967.

[73] R. Sundberg. An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Communications in Statistics: Simulation and Computation*, B5:55–64, 1976.

[74] A. J. Tsoi, H. Yang, and Shu-Ngai Yeung. European option pricing when the risk-free interest rate follows a jump process. *Communication Statistics - Stochastic models*, 16(1):143–166, 2000.

[75] J. B. Wiggins. Options values under stochstic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19:351–372, 1987.

[76] W. M. Wonham. Some applications of stochastic differential equations to optimal nonlinear filering. *Journal of Soccietу for Industrial and Applied Mathematics Series A, Control*, 2(3):347–369, 1965.

[77] M. Zakai. On the optimal filtering of diffusion process. *Z. Wahr. Verw. Gebiete*, 11:230–243, 1969.

[78] Y. Zeng. A mico-movement model of prices with Bayes estimation via filtering equation, July 2002. The conference on filtering theory and applications, University of Alberta, Canada.