

THE UNIVERSITY OF CALGARY

Stochastic Models for Electricity Prices

in Alberta

by

Lei Xiong

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

SEPTEMBER, 2004

© Lei Xiong 2004

THE UNIVERSITY OF CALGARY
FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled “Stochastic Models for Electricity Prices in Alberta” submitted by Lei Xiong in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

Dr. L.P. Bos
Department of Mathematics and
Statistics

Dr. A.F. Ware
Department of Mathematics and
Statistics

Dr. G. Sick
Haskayne School of Business

Date

Abstract

This thesis investigates the modeling of electricity prices in the Canadian province of Alberta. We model the electricity price processes as affine jump-diffusion processes, and we are able to exploit the transform analysis of Duffie, Pan and Singleton (1996) to develop computationally tractable and asymptotically efficient estimators of the parameters. We examine six mean-reverting jump-diffusion models for modeling electricity spot prices. The models which we propose have the features of multiple types of jumps, or time-varying mean and stochastic volatility. The estimation methodologies we adopt include maximum likelihood estimation based on conditional characteristic function and spectral generalized method of moments. Extensive empirical comparisons have been conducted via these estimation methods based on actual spot hourly electricity prices in Alberta.

Acknowledgments

I am indebted to many individuals for invaluable help and advice. While some are acknowledged in the text, these do not convey the extent of the assistance given to me.

First and foremost I would like to thank Professor Tony Ware for his support, guidance and patience. I learned so much from him about stochastic modeling and its application to finance. His guidance was invaluable in this thesis.

I am extremely grateful to Professor Len Bos for his incredible patience and boundless encouragement. His comments and helpful suggestions allowed me to greatly improve my work.

I also wish to acknowledge Professor Gordon Sick for his comments and corrections, which allow this thesis more satisfactory.

This thesis evolved out of a project sponsored and partially funded by NEXEN and by the Networks of Centers of Excellence for the Mathematics of Information Technology and Complex Systems (MITACS). I would also like to thank Professor Ali Lari-Lavassani for the opportunity to be involved in this project and for his generous support.

A special thanks goes to Professor Peter Zvengrowski, always friendly and helpful in editing my thesis. I am also grateful to my collaborator Zhiyong Xu for his sincere assistance in the project.

My deepest appreciation goes to my parents. They have always provided generously, not only financially, but also with their unconditional and endless love.

Finally, and by far most importantly, I thank my husband Rui. Without his

constant encouragement and love I would never have made it to this point. His silent but unwavering support has given me the strength to overcome the difficult hurdles along the way.

Table of Contents

| | |
|---|-----------|
| Approval Page | ii |
| Abstract | iii |
| Acknowledgments | iv |
| Table of Contents | vi |
| 1 Introduction | 1 |
| 2 Background | 6 |
| 2.1 The Alberta Electricity Market | 6 |
| 2.2 Models | 16 |
| 2.2.1 Geometric Brownian Motion Process | 16 |
| 2.2.2 Ornstein-Uhlenbeck Process | 17 |
| 2.2.3 Jump-diffusion model | 18 |
| 2.2.4 GARCH | 21 |
| 2.2.5 Extensions to Models | 22 |
| 3 Stochastic Models for Electricity Prices | 26 |
| 3.1 Affine Jump-Diffusion Process | 26 |
| 3.2 Model 1a | 29 |
| 3.3 Model 1b | 31 |
| 3.4 Model 2a | 33 |
| 3.5 Model 2b | 35 |
| 3.6 Model 3a | 35 |
| 3.7 Model 3b | 38 |
| 4 Parameter Estimation | 40 |
| 4.1 Introduction | 40 |
| 4.2 ML-CCF Estimators | 42 |
| 4.3 ML-MCCF Estimators | 44 |
| 4.4 Spectral GMM Estimators | 48 |
| 5 Model Comparison | 56 |
| 5.1 Data Description | 56 |
| 5.2 Calibration | 60 |
| 5.3 Goodness of Fit | 76 |

| | | |
|----------|---|------------|
| 5.4 | Robustness Test | 80 |
| 5.5 | Descriptive Statistics of Empirical vs. Calibrated Hourly Returns . . | 82 |
| 5.6 | Quantile and Quantile plot | 86 |
| 5.7 | Simulation Study | 88 |
| 5.8 | Conclusion | 90 |
| 6 | Conclusion | 108 |
| | Bibliography | 111 |

List of Tables

| | | |
|------|--|----|
| 5.1 | Descriptive statistics of Hourly EP | 57 |
| 5.2 | Descriptive statistics of Peak EP | 57 |
| 5.3 | Descriptive statistics of Offpeak EP | 58 |
| 5.4 | Descriptive statistics of deseasonalized Hourly EP | 60 |
| 5.5 | Hourly EP parameters values for Model 1a | 65 |
| 5.6 | Hourly EP parameters values for Model 2a | 67 |
| 5.7 | Hourly EP parameters values for Model 1b | 67 |
| 5.8 | Hourly EP parameters values for Model 2b | 70 |
| 5.9 | Hourly EP parameters values for Model 3a | 72 |
| 5.10 | Hourly EP parameters values for Model 3b | 74 |
| 5.11 | Bias for Model 3a | 76 |
| 5.12 | Bias for Model 3b | 77 |
| 5.13 | Hourly EP parameters values for Model 3a ($\rho = -1$) | 78 |
| 5.14 | Hourly EP parameters values for Model 3b ($\rho = -1$) | 78 |
| 5.15 | Likelihood ratio statistics fitting on Hourly EP | 79 |
| 5.16 | Schwarz criteria statistics for Models | 80 |
| 5.17 | Robustness test for Model 1a | 81 |
| 5.18 | Robustness test for Model 1b | 81 |
| 5.19 | Robustness test for Model 2a | 82 |
| 5.20 | Robustness test for Model 2b | 82 |
| 5.21 | Empirical results vs. theoretical results for Model 1a | 85 |
| 5.22 | Empirical results vs. theoretical results for Model 2a | 85 |
| 5.23 | Empirical results vs. theoretical results for Model 1b | 86 |
| 5.24 | Empirical results vs. theoretical results for Model 2b | 86 |
| 5.25 | Descriptive statistics of Peak EP and simulated paths | 90 |
| 5.26 | Descriptive statistics of Peak EP and simulated paths | 91 |

List of Figures

| | | |
|------|--|-----|
| 2.1 | Plot of the hourly electricity prices | 8 |
| 2.2 | Empirical histogram of log returns with normal density superimposed | 9 |
| 2.3 | Sample plots of the on/off-peak prices together with the histogram of log returns | 11 |
| 2.4 | Average weekday hourly prices by season | 13 |
| 2.5 | A sample of hourly electricity prices for two weeks | 14 |
| 2.6 | The moving average electricity prices vs. their variance | 15 |
| | | |
| 5.1 | Histogram of the log returns of Hourly EP | 58 |
| 5.2 | Histogram of the changes of the deseasonalized data (dX) | 59 |
| 5.3 | Results from optimization (Model 1a) | 66 |
| 5.4 | Results from optimization (Model 2a) | 68 |
| 5.5 | Results from optimization (Model 1b) | 69 |
| 5.6 | Results from optimization (Model 2b) | 71 |
| 5.7 | Results from optimization (Model 3a) | 73 |
| 5.8 | Results from optimization (Model 3b) | 75 |
| 5.9 | QQ Plots | 89 |
| 5.10 | Hourly EP superimposed by simulated paths (Model 1a) | 92 |
| 5.11 | Hourly EP superimposed by simulated paths (Model 1b) | 93 |
| 5.12 | Hourly EP superimposed by simulated paths (Model 2a) | 94 |
| 5.13 | Hourly EP superimposed by simulated paths (Model 2b) | 95 |
| 5.14 | Comparison of simulated price processes with Hourly EP (Model 1a) | 96 |
| 5.15 | Comparison of simulated price processes with Hourly EP (Model 1b) | 97 |
| 5.16 | Comparison of simulated price processes with Hourly EP (Model 2a) | 98 |
| 5.17 | Comparison of simulated price processes with Hourly EP (Model 2b) | 99 |
| 5.18 | Peak EP superimposed by simulated paths (Model 1a) | 100 |
| 5.19 | Peak EP superimposed by simulation paths (Model 1b) | 101 |
| 5.20 | Offpeak EP superimposed by simulation paths (Model 1a) | 102 |
| 5.21 | Offpeak EP superimposed by simulation paths (Model 1b) | 103 |
| 5.22 | Comparison of simulated price processes with Peak EP (Model 1a) . . | 104 |
| 5.23 | Comparison of simulated price processes with Peak EP (Model 1b) . | 105 |
| 5.24 | Comparison of simulated price processes with Offpeak EP (Model 1a) | 106 |
| 5.25 | Comparison of simulated price processes with Offpeak EP (Model 1b) | 107 |

Chapter 1

Introduction

Over the last decade, the structure of the North America electricity markets has undergone radical changes. Electricity markets are experiencing rapid deregulation and restructuring, especially with regard to generation and supply. Electricity prices are no longer controlled by regulators and now are essentially determined according to the economic rule of supply and demand. At the same time, the deregulation and restructuring of electricity markets has paved the way for a considerable amount of trading activity, which provides utilities with new opportunities but more competition as well. The resulting spot prices are crucial for the valuation of physical assets, derivatives and more generally for the risk management of utilities. The recent experience of the state of California has illuminated the importance of understanding electricity price behaviour (see [1]). While there has been a significant amount of research on commodity prices, however, the empirical studies on electricity prices are not thoroughly developed (see [26]). It is a very challenging task due to the erratic nature of electricity prices. Furthermore, most studies have been focused on the U.K. and U.S. electricity markets.

In this thesis, we will primarily study the Alberta Electricity Market. Instead of using the expected future value of electricity price as the underlying process, it seems that it is better to use electricity spot prices directly to study the dynamics of the electricity market. Introducing a convenience yield, a non-observable risk factor, as discussed in Eydeland and Hélyette (1998) (see [2]), does not really make

sense for electricity: since there is no available technique to store power (except for hydro), holding the underlying asset does not help us. Moreover, in certain regions, electricity spot markets are relatively more liquid than the corresponding futures markets, especially for electricity futures with maturity beyond 12 months (see [3]). The spot price process by itself should contain most of the fundamental properties of electricity (see [2]). We represent electricity spot prices in a natural logarithmic scale which stabilizes the statistical estimation procedure and ensures strict positivity of sample paths.¹

We begin by examining the empirical behaviour of electricity spot prices in the Alberta Electricity Market in Chapter 2. Through this examination, we believe that the class of mean-reverting jump-diffusion models are one of the ideal models to capture the characteristics of electricity prices in the Alberta Electricity Market. Some typical models that have been used to explain the dynamics of electricity prices are also discussed in Chapter 2.

The asset price models we consider all fall into the class of *affine processes*, which are given a general treatment in [5]. In Chapter 3, we outline the relevant features of this affine framework. As discussed by various authors (see [6], [7], [8]), *affine processes* can be applied to multiple types of processes such as those with jumps, time-varying mean, and stochastic volatility and even more general structures like latent variables or time-varying jumps without sacrificing computational tractability. Therefore, *affine processes* have great advantages for econometric work.

We will start with a one factor model based on the logarithm of the spot prices. This model will be considered and extended in subsequent sections of the thesis.

¹Up to now, negative electricity prices have been rarely observed, see [4].

We will deal with a one-jump version of the process and a two-jump version of the process. The one-jump version was first proposed by Das and Foresi (see [9]). The jump component has an exponentially distributed absolute value of jump size with sign of jump determined by a Bernoulli variable, while the two-jump version assumes that there are asymmetric upward jumps and downward jumps. This type of process was first considered in Duffie, Pan and Singleton (1998) and adopted by Shejue, Deng (1998) and Villaplana (2003) to mimic the “spikes” in the electricity price process (see [5], [10], [11]).

We also have two different specifications for the drift term:

- Deterministic long-term mean,
- Time-varying long-term mean incorporating on-peak and off-peak effects.

Furthermore, we also have two different specifications for the diffusion term:

- Deterministic volatility,
- Stochastic volatility, which is itself a mean-reverting square root process.

Specifically, the following alternative mean-reverting jump-diffusion (MRJD) models have been explored.

- Model 1a: one-jump version of MRJD,
- Model 1b: two-jump version of MRJD,
- Model 2a: one-jump version of MRJD with time-varying long-term mean,
- Model 2b: two-jump version of MRJD with time-varying long-term mean,

- Model 3a: one-jump version of MRJD with stochastic volatility,
- Model 3b: two-jump version of MRJD with stochastic volatility.

Affine jump-diffusion processes have analytic solutions for the conditional characteristic function associated with the conditional density function. This implies that one can obtain the conditional densities via Fourier inversion or other methods. The computations of those conditional characteristic functions (CCF) are introduced in Chapter 3.

Chapter 4 exploits the CCF to develop computationally tractable estimators of the parameters. Specifically, we use the Fourier transform of the CCF to derive the conditional log-likelihood function. When all the state variables are observable, we can use the standard maximum likelihood (ML) approach by maximizing this likelihood function. This yields the asymptotically efficient ML-CCF estimator (see [12]). But the computation cost will grow rapidly as the number of state variables increases.

When there are unobservable state variables (latent variables), the ML-CCF estimator cannot be constructed directly. Following Chacko and Viceira (1999) (see [13]), we obtain a so-called marginal CCF by integrating the latent variable out of the CCF. Based on this marginal conditional characteristic function (MCCF), we can carry out the ML-MCCF estimation. This MCCF can be used, as well, to derive closed-form expressions for the so-called conditional moments. One can construct spectral generalized methods of moments (SGMM) estimators introduced by Chacko and Viceira (1999) (see [13]). But the estimators based on the MCCF are less efficient compared to other parameter estimation methods that have been used in stochastic

volatility models. It is a trade-off between computational ease and efficiency.

Also, extensive empirical comparisons among the models have been conducted in Chapter 5. The data series we used are the hourly electricity price series for the province of Alberta from January 1, 2002 to March 31, 2004.

Chapter 2

Background

2.1 The Alberta Electricity Market

As pointed out by Daniel *et al.* [14], “Alberta is somewhat unique in Canada in that there has never existed a single, vertically integrated Crown (that is, government-owned) monopoly serving the electricity needs of the province.” Generation and retail services are open to competition, which strongly resembles the U.S. industry. However, transmission and distribution services are still under government regulation, which can be construed as a “Canadian” characteristic. The Alberta Electricity Market is relatively small. It has about 12,400 MW of installed generating capacity, including 11,500 MW in the province’s integrated electrical system and access to 950 MW from B.C. and Saskatchewan. Thermal sources account for the majority of Alberta’s installed generating capacity. The remainder is hydro and wind (see [15]). Restructuring of the electricity industry was first broached in Alberta in the early 1990s and deregulation started from January 1, 1996. Alberta has benefitted from more than 3,000 MW of new generation since 1998, an increase of about 30% in capacity. The experience of the Alberta Electricity Market reflects many aspects involved in deregulation and restructuring, and provides insights to other electricity markets.

Power generated in Alberta is exchanged through the Alberta Electric System

Operator (AESO).¹ The AESO is also responsible for dispatching all electric power generation in Alberta and directing the operation of Alberta’s electricity network to ensure reliable and economic systems. The Alberta Energy and Utilities Board (EUB), provides governance and direction of the AESO.

The System Coordination Centre is the heart of Alberta’s wholesale real-time electricity market. The System Controllers are responsible for the real-time operations in this market. They dispatch electricity to meet demand and monitor the status of the electric system through the Energy Management System (EMS). The Energy Trading System (ETS), which enables both spot and forwards electricity markets, facilitates the real-time wholesale electricity market.

All trading of power through the AESO is initiated by a process of offers and bids. Then the AESO establishes a “merit order”² to meet forecast pool demand by ranking offers and bids from lowest cost to highest cost for each hour of the day. The System Controllers dispatch offers and bids to keep the balance of supply and demand in the “merit order” and ensure the lowest cost. The last bid or offer that is dispatched every minute sets the System Marginal Price (SMP). A time-weighted average of the marginal prices at the end of the hour is calculated as the market price for that hour. There are no differential transmission charges for locations in Alberta (see [16]).

Figure 2.1 illustrates an example of hourly prices, which are measured in dollars per megawatt hour (\$/MWH, Canadian dollars being the monetary unit here and in the remainder of this thesis), during the period January 1, 2002 to March 31, 2004

¹The Alberta Electric System Operator brings together two former entities: the Power Pool of Alberta and the Transmission Administrator of Alberta.

²For more details, visit www.aeso.ca.

(19,704 observations). Figure 2.2 presents an empirical histogram for the log returns

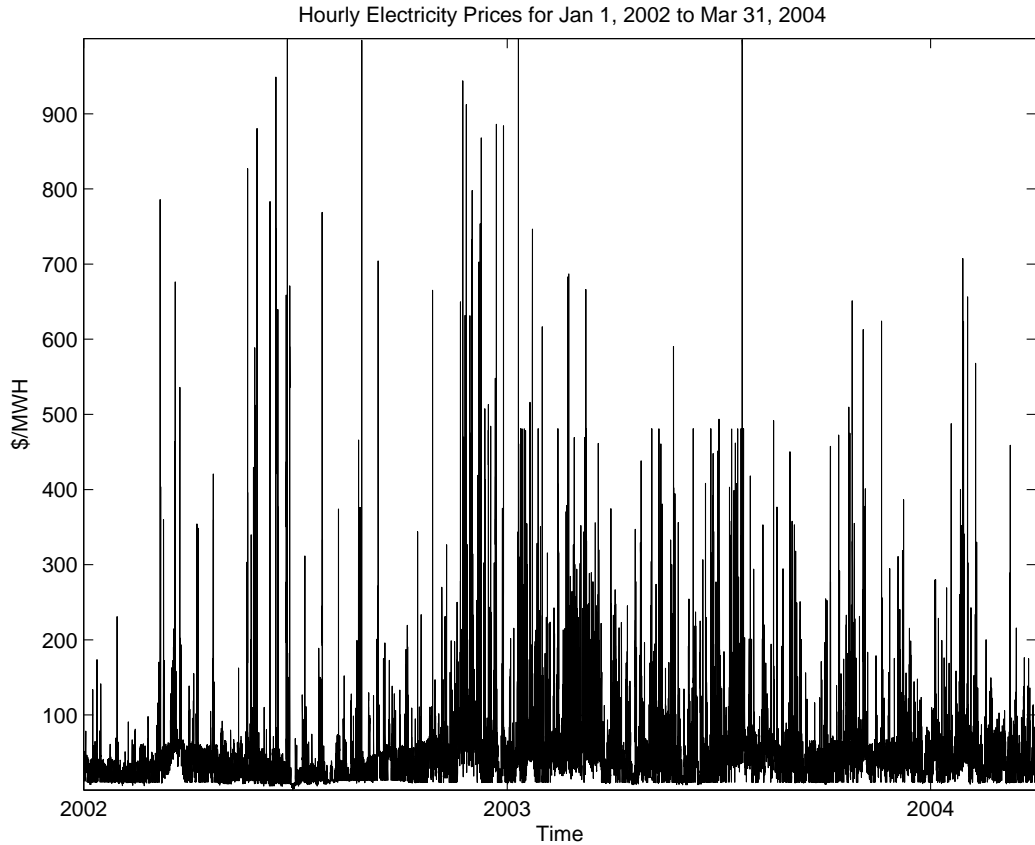


Figure 2.1: Plot of the hourly electricity prices

of the hourly electricity prices over the same period overlaid with a normal density curve.³ Clearly, the probability density curve of the change is not normal, because it has slimmer tails and a higher peak at the mean than the normal distribution.

The day time is divided into on-peak periods and off-peak periods to assist firms in managing electricity consumption and in taking advantage of lower rates in off-

³The log returns of the price process P at time t in this thesis is defined as $d(X_t) = X_{t+1} - X_t$ with $X_t = \log P_t$.

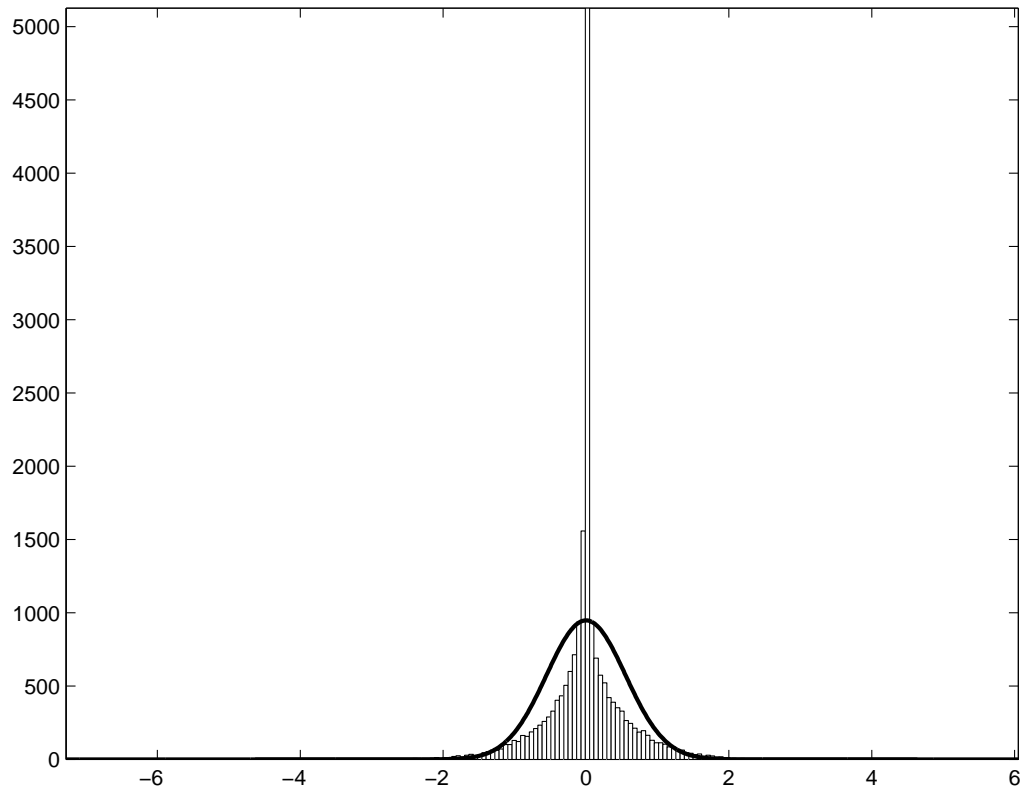


Figure 2.2: Empirical histogram of log returns with normal density superimposed
Note: the histogram is the log returns of the hourly electricity prices, the overlaid solid line is the plot of a normal density with the same mean and variance as the log returns.

peak hours, which will in turn help conserve energy and enhance available supply. According to the Alberta Energy and Utilities Board, the on-peak period is from 08:00 to 21:00 Monday through Friday inclusive, with the exception of statutory holidays.⁴ The remaining hours from Monday through Sunday are off-peak hours. Figure 2.3 includes sample plots of the on-peak and off-peak daily average electricity prices from January 1, 2002 to March 31, 2004 together with a histogram of the log returns superimposed by a normal density.

Just as electricity prices in other markets, electricity prices in the Alberta Market also display the following distinct characteristics (see [17]):

1. Pronounced cyclical effects. Electricity exhibits the most complicated cyclical patterns of all energy commodities. Normally, there are peaks reflecting heating and cooling needs. Additionally, they have hourly, daily, intradaily, and weekly patterns. Figure 2.4 presents the normalized average weekday hourly electricity prices for each season of the year 2000 to the year 2003.⁵ It may appear that temperatures are not a very significant factor in Alberta since it is hard to tell that electricity prices have strongly significant different behaviour, especially from about 8:00 p.m. to 8:00 a.m., in different seasons in Alberta from the year 2000 to the year 2003. Notice that within a time span of 24 hours, prices increase as demand increases with a distinct hourly pattern. It increases at around 5:00 a.m., as people wake up and the work hours begin. Prices continuously increase throughout the day as demand increases, peaking at around

⁴The statutory holidays in Alberta are New Year's Day, Family Day (3rd Monday of February), Good Friday, Victoria Day, Canada Day, Labour Day, Thanksgiving Day, Remembrance Day, Christmas Day.

⁵In normalizing, we adjust each average weekday hourly price by dividing by the year's average hourly price.

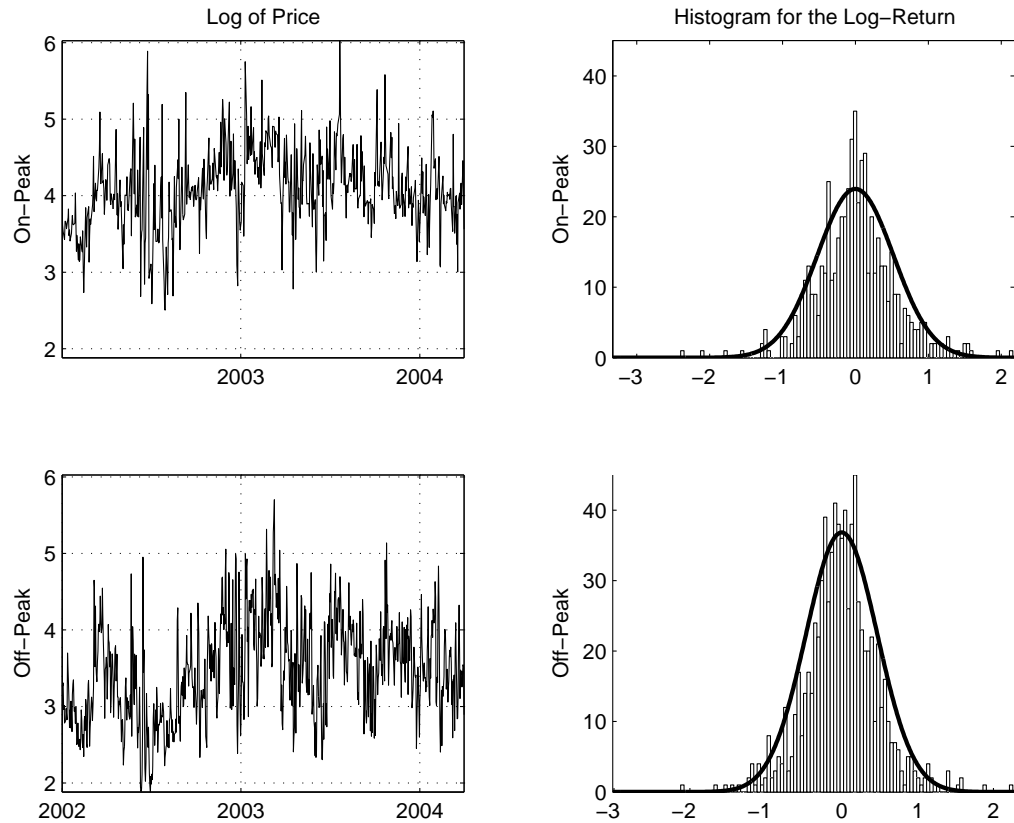


Figure 2.3: Sample plots of the on/off-peak prices together with the histogram of log returns

Note: the upper left plot is the sample plot of the logarithm of the on-peak daily average prices. The upper right plot is the histogram for the log returns of the on-peak daily average. Similarly, the lower left plot is the sample plot of the logarithm of the off-peak daily average prices. The lower right plot is the histogram for the log returns of the off-peak daily average. Both histograms illustrate the deviation from normality. Notice that, the differences between the log of the on-peak average prices and the log of the off-peak are not significant.

5:00 p.m.. After the work hours, demand shifts to primarily residential usage and prices begin to decrease. Figure 2.5 illustrates more clearly the daily usage pattern.⁶ The prices are measured in dollars per megawatt hour (\$/MWH) on the left vertical axis and demand is measured in megawatts (MW) on the right vertical axis. Moreover, it is apparent that prices, in some sense, are mimicking demand. Electricity prices are normally higher when demand is greater.

2. Sizable price “spikes”. Electricity is non-storable (other than hydro) and is more affected by transmission constraints, seasonality and weather. As shocks in demand and supply cannot be smoothed, electricity spot prices are extremely volatile and occasionally reach extremely high levels, commonly known as “spikes”. This is the most dramatic feature of Figure 2.1. We see one (or several) upward jump shortly followed by a steep downward move. It is not uncommon to see that the fluctuation of electricity prices is more than \$700/MWH. One consequence of “spikes” is the presence of so-called “fat tails”, i.e., the probability of a very large positive or negative change (though small) is much larger than permitted by a normal distribution. This can be measured by kurtosis.
3. Mean reversion. This has been well documented as an important characteristic of electricity prices. It is also quite clear in Figure 2.5. The price oscillates around the mean level and gets pulled back to this mean level rapidly after a spike. This phenomenon in particular gives rise to a high mean reversion rate.
4. Price-dependent variance. There is empirical evidence suggesting the fact that

⁶This is similar to the investigation of the hourly electricity prices from California, see [26].

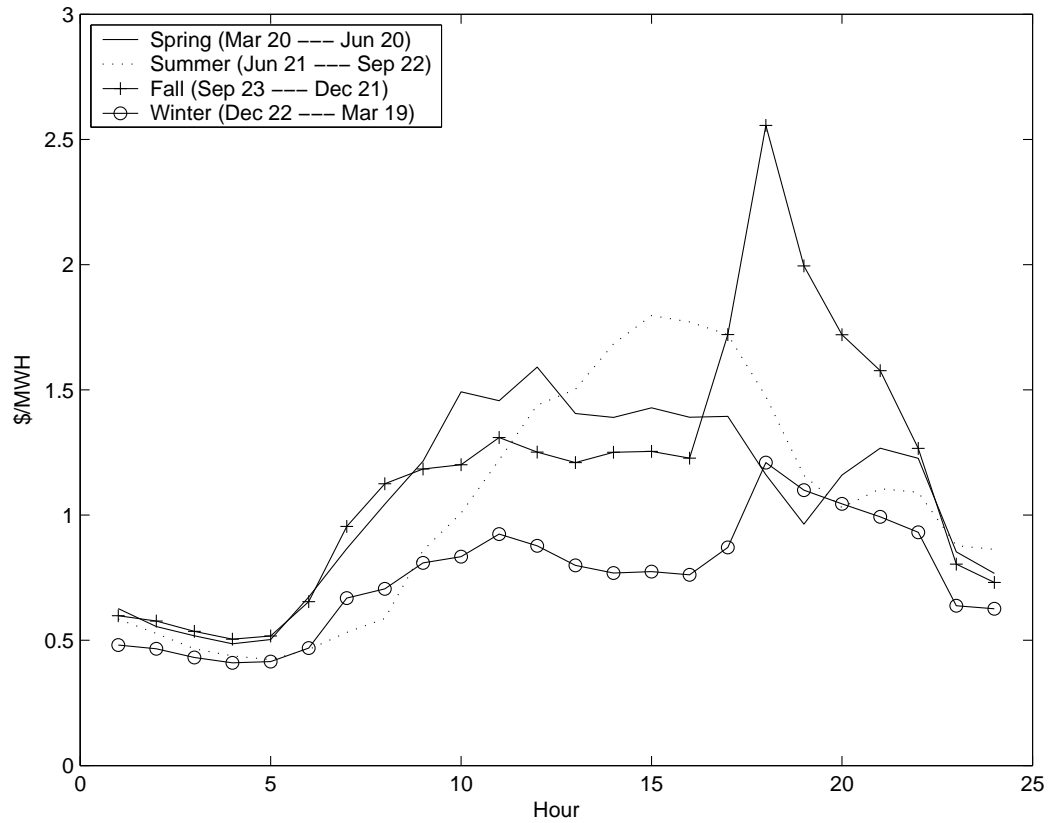


Figure 2.4: Average weekday hourly prices by season

Note: from this plot, it is hard to tell that electricity prices have strongly significant different behaviour, especially from about 8:00 p.m. to 8:00 a.m., in different seasons in Alberta from the year 2000 to the year 2003.

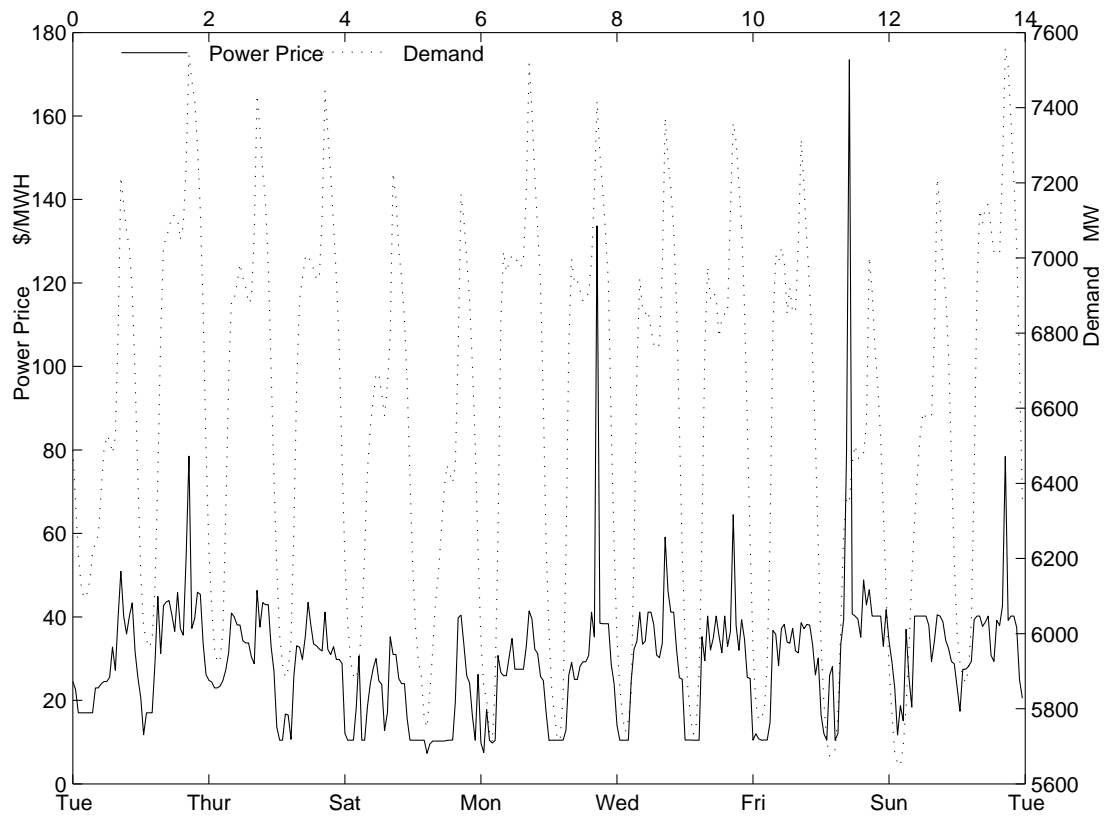


Figure 2.5: A sample of hourly electricity prices for two weeks

Note: this is a plot of sample hourly electricity prices together with the corresponding demand for two weeks.

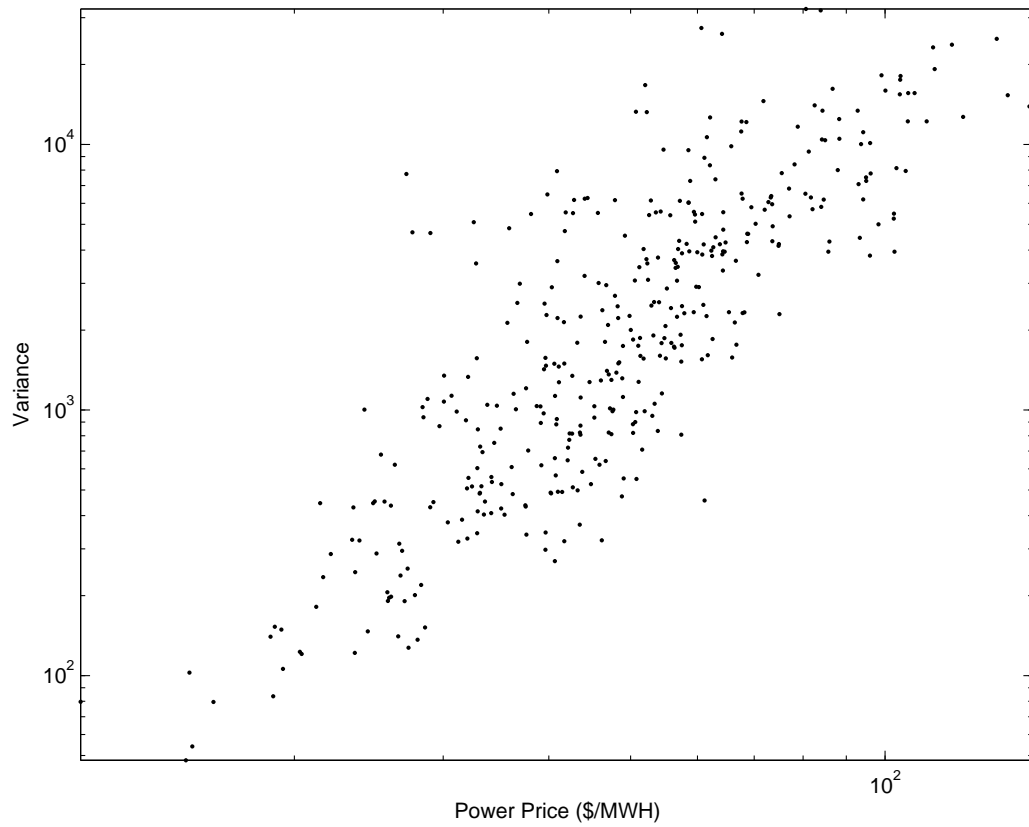


Figure 2.6: The moving average electricity prices vs. their variance
Note: this is a log-log scale plot of the 168-hour moving average electricity prices versus the variance of the corresponding 168-hour electricity prices. Electricity prices are roughly proportional to the square root of their variance.

the volatility of electricity prices is high when the aggregate demand is high and vice versa.⁷ Figure 2.6 plots the 168-hour moving average prices⁸ from Jan 1, 2002 to Mar 31, 2004 versus the variance of the corresponding 168-hour electricity prices. The prices are measured in dollars per megawatt hour (\$/MWH) on the horizontal axis. This plot clearly exhibits an upward trend, which is significant, indicating that the variance is price-dependent.

5. Non-negativity. As it costs money to produce electricity, electricity prices are normally positive although sometimes they are close to zero, as shown in Figure 2.1.

2.2 Models

Common models of asset prices offer a poor representation and forecasting of the electricity price process because they fail to capture the erratic nature of electricity prices (see [10], [17], [20]). Thus, modeling the price behaviour of electricity is a very challenging task for researchers and practitioners. Some typical models that have been used to explain the dynamics of electricity prices are discussed in the following.

2.2.1 Geometric Brownian Motion Process

The geometric Brownian motion process forms the basis for the “Black-Scholes-Merton” option pricing model and can be written as the following stochastic differ-

⁷In general, volatility is a statistical measure of the tendency of a market or security to rise or fall sharply within a short period of time. Volatility is typically calculated by using variance or annualized standard deviation of the prices or log returns (see [19]).

⁸For the spot prices $P_t, t = 1, \dots, N$, the 168-hour moving average price $M_t = \sum_{i=0}^{167} P_{t+i}/168, t = 1, \dots, N - 168$, and $\sigma_t^2 = \text{var}(\{P_{t+i}\}_{i=0}^{167})$.

ential equation (SDE):

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \quad (2.1)$$

Here dS_t is the stochastic increment over an infinitesimal time interval dt . The long-term mean μ and the diffusion coefficient σ are unknown constants, and dW_t represents an increment to a standard Brownian motion W_t which has zero mean and variance dt . This model assumes that prices are log normal, or alternatively, the log returns are normally distributed. It is relatively simple to calibrate by maximum-likelihood-based procedures and very popular in non-energy markets (see [21]). Likaa (2001) studied this model for the Alberta Electricity Market and found that it failed to predict accurately the prices of electricity due to the lack of liquidity in the Alberta Electricity Market (see [18]).

2.2.2 Ornstein-Uhlenbeck Process

Lucia and Schwartz (2000) examined the Nordpool market in terms of a totally predictable deterministic component and an Ornstein-Uhlenbeck process based on spot prices or log spot prices (see [22]). If the deterministic part is constant, then their model is a constant-mean reverting process. Mean-reverting diffusion models have been widely adapted to model financial time series, beginning with Vasicek (1977). Likaa (2001) also applied these models to the Alberta Electricity Market (see [18]).

One typical example of this type of continuous time models specifies prices as:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t, \quad (2.2)$$

where X_t is the logarithm of electricity price at time t with an initial condition X_0 ;

κ is the mean reversion rate which decides how fast the processes go back to the long-term mean level α . If the spot price is higher than e^α , then the drift, $\kappa(\alpha - X_t)$, will be negative. This brings the process back towards the long-term mean level, and similarly if the spot price is lower than e^α . The calibration of this model is also relatively easy by maximum-likelihood based procedures (see [21], [17], [23]).

If we write equation (2.2) in integral form as follows

$$X_t = e^{-\kappa t} X_0 + \alpha(1 - e^{-\kappa t}) + \int_0^t e^{-\kappa(t-s)} \sigma dW_s, \quad (2.3)$$

we obtain a first order autoregressive model (AR(1)) in continuous time and the parameters can be estimated by linear regression analysis (see [18]).

Mean-reverting diffusion models are widely used in energy markets, but in the case of electricity, we should allow for the possibility of more upside departures than downside ones and more price “spikes” in one direction than the opposite direction, which is not captured by pure mean reverting processes (see [23]). Not surprisingly, these models failed to give a reasonable prediction of electricity prices of the Alberta Electricity Market due to their limiting assumptions (see [18]).

2.2.3 Jump-diffusion model

Deng, *et al.*(1998), Clewlow and Strickland (1999), and Barz (1999) examined a broad class of stochastic models which can be used to model the behaviour of electricity prices, and performed empirical studies based on the empirical data. They found that models with mean-reversion and jumps are particularly suitable for modeling electricity spot price processes (see [17], [24], [23]).

Actually, there are many empirical studies in the literature that demonstrate the

presence of jumps as a significant feature in the behaviour of financial time series. Consequently, jump-diffusion models arise frequently in financial literature. One famous example is Merton's (1976) option pricing model, a mixed model of a continuous Brownian motion and a discrete Poisson log-normally distributed jump. In practice, besides the normal distribution, we can also have the uniform distribution (see [25]) and the exponential distribution (see [9])—including some very close variations—as qualified jump-size choices. The jump intensity can also be extended to vary with the time of day and season (see [26]).

For most jump-diffusion models, the parameters are estimated by Maximum-Likelihood (ML) estimation. ML estimation is a powerful and general method of estimating the parameters of a stochastic process when one has an analytical form for the probability density function. Ball and Torous (1985) successfully used ML estimation to estimate parameters of jump-diffusion models for NYSE stock prices. But parameter estimation in jump-diffusion models is not as easy as it may appear to be (see [27], [28]). The major caveat is that the probability density function for jump-diffusion processes cannot be determined explicitly for most models as it depends on the timing of jumps as well as their size. Moreover, the combination of the distributions of the Brownian motion and the jump component often cannot be computed analytically (see [29], [30]).

Let us take the discretized Merton model as an example,⁹

$$\Delta X_t = \left(\alpha - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta W_t + J_t\Delta P_t, \quad (2.4)$$

⁹This model is also referred to as the Bernoulli diffusion model and is discussed by Ball and Torous (1983) (see [27]). The Poisson process is approximately modeled by a Bernoulli process. For a detailed discussion, see [27].

where ΔX_t is the change of X_t during the time interval Δt , α is the long-term mean, σ is the volatility, and W_t is the Brownian Motion. Using the symbol \sim to denote the distribution of a random variable, one has $\Delta W_t \sim \mathcal{N}(0, \Delta t)$, the jump-amplitude $J_t \sim \mathcal{N}(\mu, \delta^2)$, where $\mathcal{N}(\mu, \delta^2)$ denotes a normal distribution with mean μ and standard deviation δ . Furthermore, ΔP_t is a Bernoulli process with parameter $\lambda \Delta t$. Assuming that the two normal processes are independent, then the probability density function for ΔX_t (which is denoted as $f(Y_t; \theta)$, where $Y_t = \Delta X_t$, and parameter space $\Theta = \{(\mu_1, \sigma_1, \mu_2, \sigma_2, \lambda)\}$ in the Merton model) turns out to have the form:¹⁰

$$f(Y_t; \theta) = (2\pi)^{-\frac{1}{2}} \left(a(\sigma_1^{-1} \exp\{-(Y_t - \mu_1)^2/2\sigma_1^2\}) + (1 - a)(\sigma_2^{-1} \exp\{-(Y_t - \mu_2)^2/2\sigma_2^2\}) \right), \quad \theta \in \Theta, \quad (2.5)$$

where $a = (1 - \lambda \Delta t)$, $\mu_1 = (\alpha - \frac{1}{2}\sigma^2)\Delta t$, $\sigma_1^2 = \sigma^2 \Delta t$, $\mu_2 = (\alpha - \frac{1}{2}\sigma^2)\Delta t + \mu$, and $\sigma_2^2 = \sigma^2 \Delta t + \delta^2$.

Suppose we have a sequence of T observations of X_t sampled at time t , $t = 1, \dots, T$, so that $\Delta t = 1$. The joint density function $f(\cdot)$ at the sample is given by the following equation:

$$\begin{aligned} f(Y_1, \dots, Y_{T-1}; \theta) &= \prod_{t=1}^{T-1} (2\pi)^{-\frac{1}{2}} \left(a(\sigma_1^{-1} \exp\{-(Y_t - \mu_1)^2/2\sigma_1^2\}) \right. \\ &\quad \left. + (1 - a)(\sigma_2^{-1} \exp\{-(Y_t - \mu_2)^2/2\sigma_2^2\}) \right) \\ &= \prod_{t=1}^{T-1} f(Y_t; \theta), \quad \theta \in \Theta, \end{aligned} \quad (2.6)$$

with $Y_t = X_{t+1} - X_t$. Then the logarithm of the likelihood function, $\mathcal{L}(Y_1, \dots, Y_T; \theta)$,

¹⁰For more details, see [30].

can be written as:

$$\mathcal{L}(Y_1, \dots, Y_{T-1}; \theta) = \sum_{t=1}^{T-1} \ln f(Y_t, \theta). \quad (2.7)$$

However, for example, if $\mu_1 = Y_t$ for any t then $f(Y_t; \theta)$ increases without bound as σ_1 goes to zero. Moreover, f is not zero at other observations ($\mu_1 \neq Y_t$) due to the existence of the second term of f . Thus, \mathcal{L} is unbounded and we cannot use standard ML estimation to estimate the parameters (see [32]).¹¹

Therefore ML-based estimation can be only applied to a handful of them (see [31]), such as for those that fall into the class of *affine processes*, which we will give more details later on.

2.2.4 GARCH

Autoregressive Conditional Heteroskedastic (ARCH) models were first proposed by Engle (1982) and have been widely applied in many different financial areas. Later, Bollerslev (1986) developed a generalized ARCH model, or GARCH model which allows more terms in the model. GARCH models are able to capture the very important volatility clustering phenomenon which clearly exists in electricity prices.¹² In practice, the first order GARCH(1,1) model, shown in Equation (2.8) below, is often adequate to capture the volatility structure of electricity price S_t (see [35]):

$$S_t = a_0 + a_1 S_{t-1} + \mu_t, \quad (2.8)$$

where a_0 and a_1 are unknown constants. The forecast error term is $\mu_t = \sigma_t \varepsilon_t$, where ε_t is an independent random error distributed as $\mathcal{N}(0, 1)$. The conditional variance,

¹¹However, it is still possible to obtain estimates via ML-based estimation (see [33], [30]).

¹²According to [34], volatility clustering phenomenon describes the situation wherein large volatility movements are more likely to be succeeded by further large volatility movements of either sign than by small movements.

$\sigma_t^2 := \text{var}(S_t|S_{t-1})$, is given by

$$\sigma_t^2 = b_0 + b_1\mu_t^2 + b_2\sigma_{t-1}^2. \quad (2.9)$$

Here, b_0 , b_1 and b_2 are unknown constants (see [35]). Knittel and Roberts (2001) also introduced a similar approach to model electricity prices. They specified the price level as the sum of a deterministic component and a stochastic component and adopted the Exponential GARCH (EGARCH) model of Nelson (1991) (see [26]). Also, by introducing serial correlation in the error term, they got an Autoregressive Moving Average Exogenous (ARMAX) model. They also extended this ARMAX model by incorporating temperature data to capture the seasonality of electricity prices.

2.2.5 Extensions to Models

There are also some plausible extensions of the above models, three of which are now discussed.

Time-Varying Long-term Mean

As electricity prices tend to have dramatic changes in different seasons due to heating and cooling needs and in different times of day due to change in demand, they depart from the “usual” behaviour frequently and significantly, which cannot easily be modeled by a constant long-term mean. Dummy variables can be used to reflect those properties of electricity prices. One may consider a long-term mean α_t such as

$$\alpha_t = a_1\text{peak}_t + a_2\text{offpeak}_t + \sum_{i=1}^4 b_i M_t^i, \quad (2.10)$$

where a_1 , a_2 and b_i ($i = 1, \dots, 4$) are unknown constants,

$$\begin{aligned} \text{peak}_t &= \begin{cases} 1 & \text{if in on-peak periods,} \\ 0 & \text{otherwise,} \end{cases} \\ \text{offpeak}_t &= \begin{cases} 1 & \text{if in off-peak periods,} \\ 0 & \text{otherwise,} \end{cases} \\ M_t^i &= \begin{cases} 1 & \text{if belongs to the } i\text{-th season,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{2.11}$$

Dummy variables are quite intuitive and can potentially provide some necessary flexibility. But dummy variables are very sensitive to anomalies as a result of this flexibility. For high frequency data, the use of dummy variables is always treated as an approximation since the number of steps and the placing of each step point are arbitrarily fixed (see [22]).

We may also consider including a deterministic general trend in the long-term mean. For example, one can apply a linear time trend to the logarithm of price. This implies an exponential trend for the price itself. This gives the “trending Ornstein-Uhlenbeck process” proposed by Lo and Wang (1995) (see [22]). Also, as suggested by Pilipovic (1998), a sinusoidal function can be used to capture the seasonal pattern in the price (see [20]). Moreover, we can model the long-term mean itself as a stochastic process.

Stochastic Volatility

Due to the characteristics of electricity price behaviour, some authors have argued that models for electricity prices should incorporate a form of volatility which evolves stochastically over time (see [26], [10]). One typical example is to specify the

volatility as a square-root process:

$$dv_t = \kappa(\mu - v_t)dt + \sigma\sqrt{v_t}dW_v, \quad (2.12)$$

where W_v is correlated with the Brownian motion in the spot prices. This is different from the GARCH type models (2.9), which specify the volatility as a deterministic function of lagged squared forecast error and lagged conditional variance.

It is also possible to specify the volatility as a regime-switching (see [36]) or jump-diffusion process (see [5], [10]). Estimation of stochastic volatility models presents intriguing challenges, and a variety of procedures have been proposed for fitting the models. Examples include the Bayesian Monte Carlo Markov Chain (MCMC), Efficient Method of Moments (EMM), Generalized Method of Moments (GMM), Simulated Method of Moments (SMM), and Kalman filtering methods. Two excellent recent surveys are Ghysels et al. (1995) and Shephard (1995) (see [37], [38]).

Markov Regime Switching Process

Markov regime-switching processes have been proved to be quite useful in modeling a range of financial time series including stocks, exchange rates and interest rates. Some authors also incorporated them into stochastic processes to capture the dynamics of electricity prices. For example, Hélyette and Andrea (2002) studied the dynamics of electricity prices in the major U.S. electricity markets with a combination of a deterministic term and a regime switching process. According to their study, a regime switching model ensures the Markov property in the dynamics of electricity prices, which makes the calibration and forecasting easier (see [4]). Shijie Deng (1999), Elliott, Sick and Stein (2000) and Geman (2001) also used regime switching processes to model electricity prices (see [10], [16], [4]).

The basic framework of a two-state regime switching model is to assume that there are a “volatile” state and a “normal” state of the world. The price processes behave differently depending on the state of the world. Unexpected events such as earnings announcements, scandals, or changes in macroeconomic variables signal to investors new information which often result in price processes following completely different dynamics, i.e., to switch regimes. Some plausible scenarios for electricity prices would be the forced outages of power generation plants or unexpected contingencies in transmission networks and the like.

Chapter 3

Stochastic Models for Electricity Prices

A broad class of mean-reverting jump-diffusion models will be studied in this chapter for electricity spot prices modeling. We will impose an affine structure on the coefficients of the processes, which leads to closed-form or nearly closed-form expressions for the conditional characteristic functions. We will outline the relevant features of this affine framework, which are given a general treatment in Duffie, Pan and Singleton (2000) (see [5], [12]). We will illustrate how to exploit the transform analysis introduced by Duffie, Pan and Singleton (2000) to obtain the CCF in the models we adopt.

3.1 Affine Jump-Diffusion Process

Recently, considerable attention has been focused on *affine processes*. They are flexible enough to capture certain properties such as multiple jumps, time-varying long-term mean, and stochastic volatility in various forms, that occur in many financial time series without sacrificing computational tractability. Therefore, *affine processes* have been widely used to study the term structure of interest rates, the modeling of optimal dynamic portfolios and option pricing, and so on.

We follow here the presentation in Duffie, Pan and Singleton [5]. Suppose that we are given a *strong Markov process*¹ \mathbf{X} with realizations $\{\mathbf{X}_t, 0 \leq t < \infty\}$ in some

¹For the technical definition, see [39].

state space $D \subset \mathbb{R}^n$. Under certain regularity conditions,² \mathbf{X}_t uniquely solves the following stochastic differential equation (SDE) (written in integral form):

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \boldsymbol{\mu}(\mathbf{X}_s, s) ds + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_s, s) d\mathbf{W}_s + \sum_{i=1}^m Z_t^i. \quad (3.1)$$

The jump behaviour of \mathbf{X} is governed by m types of jump processes. Each jump type Z_t^i is a pure jump process with a stochastic arrival intensity $\lambda_i(\mathbf{X}_t, t)$ for some $\lambda_i : (D, t) \mapsto \mathbb{R}^n$ and jump amplitude distribution ν_t^i on \mathbb{R}^n , where ν_t^i only depends on time t . The functions $\boldsymbol{\mu} : (D, t) \mapsto \mathbb{R}^n$ and $\boldsymbol{\sigma} : (D, t) \mapsto \mathbb{R}^{n \times n}$ must satisfy certain boundedness conditions in order to guarantee that (3.1) has a unique solution.³ The random variable \mathbf{W}_s is a standard Brownian motion in \mathbb{R}^n .

The process \mathbf{X}_t , defined by (3.1), is said to be an *affine jump-diffusion process* if⁴

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{X}, t) &= \mathbf{K}_0(t) + \mathbf{K}_1(t)\mathbf{X}, \\ \boldsymbol{\sigma}(\mathbf{X}, t)\boldsymbol{\sigma}(\mathbf{X}, t)' &= \mathbf{H}_0(t) + \sum_{k=1}^n \mathbf{H}_1^{(k)}(t)\mathbf{X}_k, \\ \lambda^i(\mathbf{X}, t) &= l_0^i(t) + \mathbf{l}_1^i(t) \cdot \mathbf{X}, \end{aligned}$$

where for each $0 \leq t < \infty$, $\mathbf{K}_0(t) \in \mathbb{R}^n$, $\mathbf{K}_1(t) \in \mathbb{R}^{n \times n}$, $\mathbf{H}_0(t) \in \mathbb{R}^{n \times n}$ and is symmetric, $\mathbf{H}_1(t) \in \mathbb{R}^{n \times n \times n}$. Also, for $k = 1, \dots, n$, $\mathbf{H}_1^{(k)}(t)$, defined to be the matrix obtained by fixing the third index of $\mathbf{H}_1(t)$ to be k , is in $\mathbb{R}^{n \times n}$ and is symmetric. Moreover, \mathbf{X}_k is the k -th entry in \mathbf{X} , $l_0^i(t) \in \mathbb{R}$, $\mathbf{l}_1^i(t) \in \mathbb{R}^n$.

Notice that given an initial condition \mathbf{X}_0 , the tuple $\theta = (\mathbf{K}_0, \mathbf{K}_1, \mathbf{H}_0, \mathbf{H}_1, l_0, \mathbf{l}_1)$ can be used to determine a transform $\Psi_\theta : \mathbb{C}^n \times [0, \infty) \times [0, \infty) \times D \rightarrow \mathbb{C}$ of \mathbf{X}_T

²The details are given in [5].

³See [5] for details.

⁴For a matrix \mathbf{C} , \mathbf{C}' is the transpose of \mathbf{C} . For column vectors \mathbf{a} and \mathbf{b} , the operation $\mathbf{a} \cdot \mathbf{b}$ is the scalar product of \mathbf{a} and \mathbf{b} .

conditional on \mathbf{X}_t , $0 \leq t \leq T$, defined by

$$\Psi_\theta(\mathbf{u}, t, T, \mathbf{X}_t) = E^\theta[\exp(\mathbf{u} \cdot \mathbf{X}_T) | \mathbf{X}_t], \quad (3.2)$$

where E^θ denotes the expectation under the distribution of \mathbf{X}_T determined by θ . Duffie *et al.* [5] have proved that if we suppose $\theta = (\mathbf{K}_0, \mathbf{K}_1, \mathbf{H}_0, \mathbf{H}_1, l_0, \mathbf{l}_1)$ is “well-behaved” at (\mathbf{u}, T) ,⁵ then the transform Ψ_θ of \mathbf{X}_t , $0 \leq t \leq T$, defined by (3.2), exists and is given by:

$$\Psi_\theta(\mathbf{u}, t, T, \mathbf{X}_t) = \exp(M(\mathbf{u}, t, T) + \mathbf{N}(\mathbf{u}, t, T) \cdot \mathbf{X}_t).$$

Here $M(\cdot)$ and $\mathbf{N}(\cdot)$ satisfy the following complex-valued Riccati equations,

$$\begin{aligned} \frac{\partial M(\mathbf{u}, t, T)}{\partial t} &= -\mathcal{A}(\mathbf{N}(\mathbf{u}, t, T), t), & M(\mathbf{u}, T, T) &= 0, \\ \frac{\partial \mathbf{N}(\mathbf{u}, t, T)}{\partial t} &= -\mathcal{B}(\mathbf{N}(\mathbf{u}, t, T), t), & \mathbf{N}(\mathbf{u}, T, T) &= \mathbf{u}, \end{aligned} \quad (3.3)$$

where, for any $\mathbf{c} \in \mathbb{C}^n$,⁶

$$\begin{aligned} \mathcal{A}(\mathbf{c}, t) &= \mathbf{K}_0(t) \cdot \mathbf{c} + \frac{1}{2} \mathbf{c}' \mathbf{H}_0(t) \mathbf{c} + \sum_{i=1}^m l_0^i (\varphi_i(\mathbf{c}) - 1), \\ \mathcal{B}(\mathbf{c}, t) &= \mathbf{K}_1(t)' \mathbf{c} + \frac{1}{2} \mathbf{c}' \mathbf{H}_1(t) \mathbf{c} + \sum_{i=1}^m \mathbf{l}_1^i (\varphi_i(\mathbf{c}) - 1). \end{aligned} \quad (3.4)$$

Here $\varphi_i(\mathbf{c})$ is the so-called “jump transform” for the i -th jump. It is given by $\varphi_i(\mathbf{c}) = \int_{\mathbb{R}^n} \exp(\mathbf{c} \cdot \mathbf{z}) d\nu_t^i(\mathbf{z})$ whenever the integral is well defined. For example, for a normally distributed jump size with mean μ and variance σ^2 , it can be shown that $\varphi(c) = \exp(\mu c + \frac{1}{2} \sigma^2 c^2)$; for an exponentially distributed jump size with mean μ , $\varphi(c) = \frac{1}{1 - \mu c}$ (see [5]).

⁵See [5] for details.

⁶ \mathbb{C}^n denotes the set of n -tuples of complex numbers; \mathbb{R}^n denotes the set of n -tuples of real numbers.

By setting $\mathbf{u} = i\mathbf{s}$ ($i = \sqrt{-1}$), one can obtain the CCF of \mathbf{X}_T conditional on \mathbf{X}_t as:

$$\begin{aligned}\phi(\mathbf{s}, \theta, \mathbf{X}_T | \mathbf{X}_t) &:= \Psi_\theta(i\mathbf{s}, t, T, \mathbf{X}_t) \\ &= E^\theta[\exp(i\mathbf{s} \cdot \mathbf{X}_T) | \mathbf{X}_t] \\ &= \int_{\mathbb{R}^N} \exp(i\mathbf{s} \cdot \mathbf{X}_T) f(\mathbf{X}_T, \theta | \mathbf{X}_t) d\mathbf{X}_T,\end{aligned}\tag{3.5}$$

with $f(\mathbf{X}_T, \theta | \mathbf{X}_t)$ is the conditional density of \mathbf{X}_T conditional on \mathbf{X}_t .

Notice that the CCF is actually the Fourier transform⁷ of the conditional density. Through the inverse Fourier transform, one can recover the conditional density function from the CCF and implement a usual ML estimation. We will exploit this CCF of discretely sampled observations to develop computationally tractable estimators of parameters in Chapter 4.

3.2 Model 1a

Our first attempt to capture the mean-reversion and spikes present in electricity prices is by a standard mean-reverting jump-diffusion process. We start with specifying the logarithm of the spot price, X_t , using a model adopted from Das and Foresi (1996). The diffusion part is represented by an Ornstein-Uhlenbeck process and the jump component has exponentially distributed absolute value of jump size with the

⁷Brief summary of Fourier transform:

1. Given a function $f(x)$, we take the Fourier transform to be $\hat{f}(\xi) = \mathcal{F}[f](\xi) = \int_{-\infty}^{\infty} f(x)e^{ix\xi} dx$
2. Integration by parts reveals that the Fourier transform takes differentiation to multiplication (by ξ): $\mathcal{F}[f_x](\xi) = -i\xi\hat{f}(\xi)$
3. If $h(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy$ then $\hat{h}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$
4. The Fourier transform is invertible, and $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix\xi} \hat{f}(\xi) d\xi$ is called the inverse Fourier transform.

For more details of Fourier transform, check the link mathworld.wolfram.com/FourierTransform.html.

sign of the jump determined by a Bernoulli variable. This is encapsulated by the following SDE:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t + Q_t dP_t(\omega). \quad (3.6)$$

Here, the tuple $\theta = [\kappa, \alpha, \sigma^2, \omega, \psi, \gamma]$ gives the unknown constant parameters. Specifically, κ is the mean reversion rate, α is the long-term mean, W_t is a standard Brownian motion with $dW_t \sim \mathcal{N}(0, dt)$ for an infinitesimal time interval dt , and P_t is a discontinuous, one dimensional standard Poisson process with arrival rate ω . During dt , $dP_t = 1$ if there is a jump, $dP_t = 0$ otherwise. The jump amplitude Q_t is exponentially distributed with mean γ and the sign of the jump Q_t is distributed as a Bernoulli random variable with parameter ψ . We assume that the Brownian motion, Poisson process, and random jump amplitude are all Markov and pairwise independent.

Notice that this equation fits in the framework outlined in Section 3.1 with

$$K_0 = \kappa\alpha, \quad K_1 = -\kappa, \quad H_0 = \sigma^2, \quad H_1 = 0, \quad l_0 = \omega, \quad l_1 = 0. \quad (3.7)$$

Thus the CCF of X_T given X_t , $\phi(s, \theta, X_T|X_t)$, takes the form:

$$\begin{aligned} \phi(s, \theta, X_T|X_t) &= E[\exp(isX_T)|X_t] \\ &= \exp(A(s, t, T, \theta) + B(s, t, T, \theta)X_t), \end{aligned} \quad (3.8)$$

where $A(\cdot)$ and $B(\cdot)$ satisfy the following system of complex-valued ordinary differential equations (ODE):

$$\begin{aligned} \frac{\partial A(s, t, T, \theta)}{\partial t} &= -\kappa\alpha B(s, t, T, \theta) - \frac{1}{2}\sigma^2 B^2(s, t, T, \theta) - \omega(\varphi(B(s, t, T, \theta)) - 1), \\ \frac{\partial B(s, t, T, \theta)}{\partial t} &= \kappa B(s, t, T, \theta), \end{aligned} \quad (3.9)$$

with boundary conditions:

$$A(s, T, T, \theta) = 0, \quad B(s, T, T, \theta) = is. \quad (3.10)$$

Here, the ‘‘jump transform’’ $\varphi(B(s, t, T, \theta))$ is given by:

$$\begin{aligned} \varphi(B(s, t, T, \theta)) &= \psi \int_0^\infty \exp(B(s, t, T, \theta)z) \frac{1}{\gamma} \exp\left(-\frac{z}{\gamma}\right) dz \\ &\quad + (1 - \psi) \int_0^\infty \exp(-B(s, t, T, \theta)z) \frac{1}{\gamma} \exp\left(-\frac{z}{\gamma}\right) dz \\ &= \frac{\psi}{1 - B(s, t, T, \theta)\gamma} + \frac{1 - \psi}{1 + B(s, t, T, \theta)\gamma}. \end{aligned} \quad (3.11)$$

Now, we solve the system (3.9) for $A(\cdot)$ and $B(\cdot)$ and apply the corresponding boundary conditions to obtain (after some calculation)

$$\begin{aligned} A(s, t, T, \theta) &= i\alpha s(1 - e^{-\kappa(T-t)}) - \frac{\sigma^2 s^2}{4\kappa}(1 - e^{-2\kappa(T-t)}) \\ &\quad + \frac{i\omega(1 - 2\psi)}{\kappa} (\arctan(\gamma s e^{-\kappa(T-t)}) - \arctan(\gamma s)) \\ &\quad + \frac{\omega}{2\kappa} \ln\left(\frac{1 + \gamma^2 s^2 e^{-2\kappa(T-t)}}{1 + \gamma^2 s^2}\right), \\ B(s, t, T, \theta) &= i s e^{-\kappa(T-t)}. \end{aligned} \quad (3.12)$$

3.3 Model 1b

In this model, we allow for asymmetric upward and downward jumps (see [10], [11]), each with an exponentially distributed jump magnitude. More specifically, we suppose that the logarithm of the spot price X_t satisfies the following SDE:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t + Q_t^u dP_t^u(\omega_u) + Q_t^d dP_t^d(\omega_d). \quad (3.13)$$

Again, κ is the mean reversion rate, α is the long-term mean, and W_t is a standard Brownian motion with $dW_t \sim \mathcal{N}(0, dt)$. The jump behaviour of X_t is governed

by two types of jumps: upward jumps and downward jumps. The upward jumps Q_t^u are exponentially distributed with positive mean (γ_u) and jump arrival rate ω_u . The downward jumps (Q_t^d) are also exponentially distributed negative mean γ_d and jump arrival rate ω_d . Again, P_t^u and P_t^d are two independent discontinuous, one dimensional standard Poisson processes with arrival rate ω_u and ω_d respectively. Notice that this equation also fits in the framework outlined in Section 3.1.

Thus the transform analysis can be implemented in this case and the CCF can be written out explicitly for this model

$$\begin{aligned}\phi(s, \theta, X_T|X_t) &= E[\exp(isX_T)|X_t] \\ &= \exp(A(s, t, T, \theta) + B(s, t, T, \theta)X_t).\end{aligned}\tag{3.14}$$

Here $A(\cdot)$ and $B(\cdot)$ satisfy the complex-valued system of ODE's:

$$\begin{aligned}\frac{\partial A(s, t, T, \theta)}{\partial t} &= -\kappa\alpha B(s, t, T, \theta) - \frac{1}{2}\sigma^2 B^2(s, t, T, \theta) \\ &\quad - \omega_u(\varphi_u(B(s, t, T, \theta)) - 1) - \omega_d(\varphi_d(B(s, t, T, \theta)) - 1), \\ \frac{\partial B(s, t, T, \theta)}{\partial t} &= \kappa B(s, t, T, \theta),\end{aligned}\tag{3.15}$$

with boundary conditions:

$$A(s, T, T, \theta) = 0, \quad B(s, T, T, \theta) = is.\tag{3.16}$$

Here the ‘‘jump transform’’ for the upward jump is given by:

$$\begin{aligned}\varphi_u(B(s, t, T, \theta)) &= \int_0^\infty \exp(B(s, t, T, \theta)z) \left[\frac{1}{\gamma_u} \exp\left(-\frac{z}{\gamma_u}\right) \right] dz \\ &= \frac{1}{1 - B(s, t, T, \theta)\gamma_u}.\end{aligned}\tag{3.17}$$

Similarly, the ‘‘jump transform’’ for the downward jump is given by:

$$\varphi_d(B(s, t, T, \theta)) = \frac{1}{1 - B(s, t, T, \theta)\gamma_d}.\tag{3.18}$$

Again, after some computation we solve for $A(\cdot)$ and $B(\cdot)$ and apply the corresponding boundary conditions to get (after some calculation)

$$A(s, t, T, \theta) = i\alpha s(1 - e^{-\kappa(T-t)}) - \frac{\sigma^2 s^2}{4\kappa}(1 - e^{-2\kappa(T-t)}) + \frac{\omega_u}{\kappa} \ln\left(\frac{1 - is\gamma_u e^{-\kappa(T-t)}}{1 - is\gamma_u}\right) + \frac{\omega_d}{\kappa} \ln\left(\frac{1 - is\gamma_d e^{-\kappa(T-t)}}{1 - is\gamma_d}\right), \quad (3.19)$$

$$B(s, t, T, \theta) = ise^{-\kappa(T-t)}.$$

3.4 Model 2a

We impose upon the price process (3.6) a time-varying component in the drift by replacing the long-term mean α with a deterministic function $\alpha(t)$. The logarithm of electricity spot price is thus defined by

$$dX_t = \kappa(\alpha(t) - X_t)dt + \sigma dW_t + Q_t dP_t(\omega). \quad (3.20)$$

Recall that the seasonal effects on electricity prices are not strongly significant for the specific data we analyze (refer to Figure 2.4). We only incorporate the on-peak and off-peak effects into the price process and consider the following form for $\alpha(t)$:

$$\alpha(t) = \alpha_1 \text{peak}_t + \alpha_2 \text{offpeak}_t, \quad (3.21)$$

where

$$\begin{aligned} \text{peak}_t &= \begin{cases} 1 & \text{if in on-peak periods,} \\ 0 & \text{otherwise,} \end{cases} \\ \text{offpeak}_t &= \begin{cases} 1 & \text{if in off-peak periods,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (3.22)$$

This is also an affine process. By the same methods, the CCF of X_T given X_t can be expressed in closed-form in this case:

$$\begin{aligned}\phi(s, \theta, X_T|X_t) &= E[\exp(isX_T)|X_t] \\ &= \exp(A(s, t, T, \theta) + B(s, t, T, \theta)X_t).\end{aligned}\tag{3.23}$$

Here $A(\cdot)$ and $B(\cdot)$ satisfy the complex-valued system of ODE's:

$$\begin{aligned}\frac{\partial A(s, t, T, \theta)}{\partial t} &= -\kappa\alpha(t)B(s, t, T, \theta) - \frac{1}{2}\sigma^2 B^2(s, t, T, \theta) \\ &\quad - \omega(\varphi(B(s, t, T, \theta)) - 1), \\ \frac{\partial B(s, t, T, \theta)}{\partial t} &= \kappa B(s, t, T, \theta),\end{aligned}\tag{3.24}$$

with boundary conditions:

$$A(s, T, T, \theta) = 0, \quad B(s, T, T, \theta) = is.\tag{3.25}$$

We can obtain $A(\cdot)$ and $B(\cdot)$ in the same fashion:

$$\begin{aligned}A(s, t, T, \theta) &= L(s, t, T, \theta) - \frac{\sigma^2 s^2}{4\kappa}(1 - e^{-2\kappa(T-t)}) \\ &\quad + \frac{i\omega(1 - 2\psi)}{\kappa}(\arctan(\gamma s e^{-\kappa(T-t)}) - \arctan(\gamma s)) \\ &\quad + \frac{\omega}{2\kappa} \ln\left(\frac{1 + \gamma^2 s^2 e^{-2\kappa(T-t)}}{1 + \gamma^2 s^2}\right), \\ B(s, t, T, \theta) &= i s e^{-\kappa(T-t)},\end{aligned}\tag{3.26}$$

Here, let $t = t_0 < t_1 \cdots < t_N = T$, then we have

$$\begin{aligned}L(s, t, T, \theta) &= \int_t^T \kappa\alpha(t) i s e^{-\kappa(T-t)} dt \\ &= i s \sum_{j=1}^N \alpha e^{-\kappa(T-t_j)} (1 - e^{-\kappa(t_j - t_{j-1})}),\end{aligned}\tag{3.27}$$

where

$$\alpha = \begin{cases} \alpha_1 & \text{if } [t_{j-1}, t_j] \text{ is in on-peak periods,} \\ \alpha_2 & \text{otherwise.} \end{cases}\tag{3.28}$$

3.5 Model 2b

We consider the following extension to the model (3.13):

$$dX_t = \kappa(\alpha(t) - X_t)dt + \sigma dW_t + Q_t^u dP_t(\omega_u) + Q_t^d dP_t(\omega_d), \quad (3.29)$$

where $\alpha(t)$ is defined the same as in (3.21).

Then similar to the above models, the CCF of X_T given X_t is of the form:

$$\begin{aligned} \phi(s, \theta, X_T | X_t) &= E[\exp(isX_T) | X_t] \\ &= \exp(A(s, t, T, \theta) + B(s, t, T, \theta)X_t), \end{aligned} \quad (3.30)$$

where

$$\begin{aligned} A(s, t, T, \theta) &= L(s, t, T, \theta) - \frac{\sigma^2 s^2}{4\kappa} (1 - e^{-2\kappa(T-t)}) \\ &\quad + \frac{\omega_u}{\kappa} \ln \left(\frac{1 - is\gamma_u e^{-\kappa(T-t)}}{1 - is\gamma_u} \right) + \frac{\omega_d}{\kappa} \ln \left(\frac{1 - is\gamma_d e^{-\kappa(T-t)}}{1 - is\gamma_d} \right), \\ B(s, t, T, \theta) &= ise^{-\kappa(T-t)}. \end{aligned} \quad (3.31)$$

with $L(s, t, T, \theta)$ defined in Equation (3.27).

3.6 Model 3a

It has been well documented that jumps alone are inadequate to mimic the level of skewness present in electricity spot prices. Kaminski (1997) and Deng (1998) emphasized the need to incorporate stochastic volatility in the modeling of electricity spot prices (see [40], [10]). Also, according to the study of Mika and Andrew, volatility in electricity prices varies over time and is likely mean-reverting itself (see [41]). So we consider the following two-factor affine process (3.32) to model electricity spot

prices. Let X_t be the logarithm of the spot price of electricity and V_t be the volatility of the price process which evolves stochastically over time. We have

$$d \begin{bmatrix} X_t \\ V_t \end{bmatrix} = \begin{bmatrix} \kappa(\alpha - X_t) \\ \kappa_v(\alpha_v - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{1-\rho^2}\sqrt{V_t} & \rho\sqrt{V_t} \\ 0 & \sigma_v\sqrt{V_t} \end{bmatrix} \begin{bmatrix} dW_t \\ dW_v \end{bmatrix} + \begin{bmatrix} Q_t dP_t(\omega) \\ 0 \end{bmatrix}. \quad (3.32)$$

Here, κ is the mean reversion rate of the log prices, α is the long-term mean of the log prices, P_t is a discontinuous, one dimensional standard Poisson process with arrival rate ω , the amplitude of Q_t is exponentially distributed with mean γ , and the sign of Q_t is distributed as a Bernoulli random variable with parameter ψ . The two random variables W_t and W_v are two uncorrelated standard Brownian motions. Also, κ_v is the mean reversion rate of the volatility V_t , α_v is the long-term mean of V_t , and σ_v is the volatility of V_t .

This model also fits in the framework outlined in Section 3.1 with

$$\begin{aligned} \mathbf{K}_0 &= \begin{bmatrix} \kappa\alpha \\ \kappa_v\alpha_v \end{bmatrix}, & \mathbf{K}_1 &= \begin{bmatrix} -\kappa & 0 \\ 0 & -\kappa_v \end{bmatrix}, & H_0 &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \\ \mathbf{H}_1^{(1)} &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, & \mathbf{H}_1^{(2)} &= \begin{bmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{bmatrix}, & & (3.33) \\ l_0 &= \omega, & \mathbf{l}_1 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Suppose that the jump component in the logarithm of the spot price is defined as in

Model 1a and Model 2a. Then the “jump transform” is given by

$$\begin{aligned} \varphi \left(\begin{bmatrix} A(\cdot) \\ B(\cdot) \end{bmatrix} \right) &= \int_0^\infty \left(\psi \exp(A(\cdot)z) + (1 - \psi) \exp(-A(\cdot)z) \right) \frac{1}{\gamma} \exp\left(-\frac{z}{\gamma}\right) dz \\ &= \frac{\psi}{1 - A(\cdot)\gamma} + \frac{1 - \psi}{1 + A(\cdot)\gamma}. \end{aligned} \quad (3.34)$$

Thus the CCF is of the form

$$\begin{aligned} &\phi(s_x, s_v, \theta, X_T, V_T | X_t, V_t) \\ &= E^\theta [\exp(is_x X_T + is_v V_T) | (X_t, V_t)] \\ &= \exp(A(s_x, s_v, t, T, \theta)X_t + B(s_x, s_v, t, T, \theta)V_t + C(s_x, s_v, t, T, \theta)), \end{aligned} \quad (3.35)$$

where $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ satisfy the following complex-valued Riccati equations,

$$\begin{cases} \frac{\partial A(\cdot)}{\partial t} = \kappa A(\cdot), \\ \frac{\partial B(\cdot)}{\partial t} = \kappa_v B(\cdot) - \frac{1}{2} A(\cdot)(A(\cdot) + \rho \sigma_v B(\cdot)) \\ \quad - \frac{1}{2} B(\cdot)(A(\cdot) \rho \sigma_v + B(\cdot) \sigma_v^2), \\ \frac{\partial C(\cdot)}{\partial t} = -\kappa \alpha A(\cdot) - \kappa_v \alpha_v B(\cdot) - \omega \left(\varphi \left(\begin{bmatrix} A(\cdot) \\ B(\cdot) \end{bmatrix} \right) - 1 \right), \end{cases} \quad (3.36)$$

with boundary conditions:

$$A(s_x, s_v, T, T, \theta) = is_x, \quad B(s_x, s_v, T, T, \theta) = is_v, \quad C(s_x, s_v, T, T, \theta) = 0. \quad (3.37)$$

We can still solve the first equation for $A(\cdot)$ and apply the corresponding initial conditions to obtain

$$A(\cdot) = is_x e^{-\kappa(T-t)}. \quad (3.38)$$

However, we don't have closed forms for $B(\cdot)$ and $C(\cdot)$ and need to solve them numerically.

3.7 Model 3b

We now redefine the model (3.32) above by allowing two types of jumps just as before:

$$d \begin{bmatrix} X_t \\ V_t \end{bmatrix} = \begin{bmatrix} \kappa(\alpha - X_t) \\ \kappa_v(\alpha_v - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{1-\rho^2}\sqrt{V_t} & \rho\sqrt{V_t} \\ 0 & \sigma_v\sqrt{V_t} \end{bmatrix} \begin{bmatrix} dW_t \\ dW_v \end{bmatrix} + \begin{bmatrix} Q_t^u dP_t(\omega_u) + Q_t^d dP_t(\omega_d) \\ 0 \end{bmatrix}. \quad (3.39)$$

Then the “jump transform” for the upward jump is given by

$$\begin{aligned} \varphi_u \left(\begin{bmatrix} A(\cdot) \\ B(\cdot) \end{bmatrix} \right) &= \int_0^\infty \exp(A(\cdot)z) \left[\frac{1}{\gamma_u} \exp\left(-\frac{z}{\gamma_u}\right) \right] dz \\ &= \frac{1}{1 - A(\cdot)\gamma_u}. \end{aligned} \quad (3.40)$$

Similarly, the “jump transform” for the downward jump is given by

$$\varphi_d \left(\begin{bmatrix} A(\cdot) \\ B(\cdot) \end{bmatrix} \right) = \frac{1}{1 - A(\cdot)\gamma_d}. \quad (3.41)$$

As this model also fits in the setting of affine processes, one can obtain the CCF in the same fashion:

$$\begin{aligned} \phi(s_x, s_v, \theta, X_T, V_T | X_t, V_t) &= E^\theta[\exp(is_x X_T + is_v V_T) | (X_t, V_t)] \\ &= \exp(A(s_x, s_v, t, T, \theta)X_t + B(s_x, s_v, t, T, \theta)V_t + C(s_x, s_v, t, T, \theta)), \end{aligned} \quad (3.42)$$

where $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ satisfy the following complex-valued Riccati equations,

$$\left\{ \begin{array}{l} \frac{\partial A(\cdot)}{\partial t} = \kappa A(\cdot), \\ \frac{\partial B(\cdot)}{\partial t} = \kappa_v B(\cdot) - \frac{1}{2} A(\cdot)(A(\cdot) + \rho \sigma_v B(\cdot)) \\ \quad - \frac{1}{2} B(\cdot)(A(\cdot) \rho \sigma_v + B(\cdot) \sigma_v^2), \\ \frac{\partial C(\cdot)}{\partial t} = -\kappa \alpha A(\cdot) - \kappa_v \alpha_v B(\cdot) \\ \quad - \omega_u(\varphi_u \left(\begin{bmatrix} A(\cdot) \\ B(\cdot) \end{bmatrix} \right) - 1) - \omega_d(\varphi_d \left(\begin{bmatrix} A(\cdot) \\ B(\cdot) \end{bmatrix} \right) - 1). \end{array} \right. \quad (3.43)$$

with boundary conditions:

$$A(s_x, s_v, T, T, \theta) = i s_x, \quad B(s_x, s_v, T, T, \theta) = i s_v, \quad C(s_x, s_v, T, T, \theta) = 0. \quad (3.44)$$

Again, we obtain $A(\cdot) = i s_x e^{-\kappa(T-t)}$ and need to solve $B(\cdot)$ and $C(\cdot)$ numerically.

Chapter 4

Parameter Estimation

4.1 Introduction

Affine processes are flexible enough to allow us to capture the special characteristics of electricity prices such as mean-reversion, seasonality, and “spikes” (see [10]). Moreover, under suitable regularity conditions, one can explore the information from the CCF of discretely sampled observations to develop computationally tractable and *asymptotically efficient*¹ estimators of the parameters of affine processes (see [5]). Moreover, the CCF is unique and contains the same information as the conditional density function through the Fourier transform. We can use it to recover the conditional density function via the Fourier transform and implement a usual ML estimation. This is the approach of ML-CCF estimation, for which we will give more details later on. If the N -dimensional state variables are all observable, ML-CCF estimation can be implemented and the so obtained ML-CCF estimators are asymptotically efficient (see [12]).

But the estimation can be costly in higher dimensions ($N \geq 2$) because we need to compute the multivariate Fourier inversions repeatedly and accurately in order to maximize the likelihood function. According to Singleton (2001), considerable computational saving can be achieved by using limited-information ML-CCF (LML-

¹The ratio of the Rao-Cramér lower bound to the actual variance of any unbiased estimation of a parameter is called the efficiency of that statistic. If the efficiency tends to 1 as the number of observations increases, the estimator is said to be asymptotically efficient. For more details about Rao-Cramér bounds and efficiency, see [42].

CCF) estimation (see [12]). Suppose $\{\mathbf{X}_t, t = 1, 2, \dots\}$ is a set of discretely sampled observations of a N -dimensional state variable with a joint CCF $\phi(\mathbf{s}, \theta, \mathbf{X}_{t+1}|\mathbf{X}_t)$. Let $\boldsymbol{\eta}_j$ denote a N -dimensional selection vector where the j^{th} entry is 1 and zeros elsewhere. Define $X_{t+1}^j := \boldsymbol{\eta}_j \cdot \mathbf{X}_{t+1}$, then the conditional density of X_{t+1}^j conditioned on \mathbf{X}_t is the inverse Fourier transform of $\phi(\xi \boldsymbol{\eta}_j, \theta, \mathbf{X}_{t+1}|\mathbf{X}_t)$ with some scalar ξ :²

$$f_j(X_{t+1}^j, \theta|\mathbf{X}_t) = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(\xi \boldsymbol{\eta}_j, \theta, \mathbf{X}_{t+1}|\mathbf{X}_t) e^{-i\xi \boldsymbol{\eta}_j' \mathbf{X}_{t+1}} d\xi. \quad (4.1)$$

The basic idea behind this is to exploit the information in $f_j(X_{t+1}^j, \theta|\mathbf{X}_t)$ instead of the information in the joint conditional density function,

$$f(\mathbf{X}_{t+1}, \theta|\mathbf{X}_t) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \phi(\mathbf{s}, \theta, \mathbf{X}_{t+1}|\mathbf{X}_t) e^{-i\mathbf{s}' \mathbf{X}_{t+1}} d\mathbf{s}. \quad (4.2)$$

Thus the estimation involves at most N one-dimensional integrations instead of doing a N -dimensional integration. The estimators obtained are called LML-CCF estimators. Although the LML-CCF estimators do not exploit any information about the joint conditional density function, they are typically more efficient than the quasi-maximum likelihood (QML) estimators for affine diffusions (see [12]).³

But for those multi-factor models with unobservable (latent) state variables such as stochastic volatility models, the ML-CCF or LML-CCF estimators cannot be obtained. There are several recent papers that discuss the methodologies related to CCF-based estimators of stochastic volatility models. Singleton (1999) (see [12]) proposed a Simulated Method of Moments (SMM-CCF) estimator; Jiang and Knight (1999) (see [43]) explored the Method of System of Moments (MSM) estimators;

²Refer back to (3.5) in Chapter 3.

³QML acts as if the data were generated by a density function that provides an estimator that is easy to obtain. This method only makes assumptions about the mean and the second moment. For further details, see Bollerslev and Wooldridge (1992) and Newey and Steigerwald (1997).

Chacko and Viceira (1999) (see [13]) considered the so-called Spectral Generalized Method of Moments (SGMM) estimators. In this thesis, we adopt the idea of SGMM because this methodology is more computationally tractable than the others (see [12]). To deal with stochastic volatility models, they derived the stationary (unconditional) characteristic function⁴ from the CCF of the volatility, and utilized this CCF to obtain a so-called marginal CCF. We apply a ML type estimation based on the so-called marginal CCF (ML-MCCF) to estimate stochastic volatility models. Furthermore, we also introduce SGMM estimators based on the so-called marginal CCF to estimate stochastic volatility models.

4.2 ML-CCF Estimators

ML estimation is the most common method of estimating the parameters of stochastic processes if the probability density has an analytical form. It provides a *consistent* approach to parameter estimation problems and ML estimators become *minimum variance unbiased* estimators as the sample size increases.⁵

Suppose that \mathbf{X} is a N -dimensional continuous random variable with probability density function $f(\mathbf{X}, \theta)$ where $\theta = \{\theta_1, \dots, \theta_k\}$ are k unknown constant parameters which need to be estimated. Given a sequence of observations $\{\mathbf{X}_t\}$ sampled at

⁴It turns out that as $t \rightarrow \infty$, there exists a limiting distribution for X_t . This limiting distribution is called the stationary distribution and its Fourier transform is called the stationary characteristic function (see [44]).

⁵Any statistic that converges in probability to a parameter is called a consistent estimator of that parameter. By unbiased, we mean that the mathematical expectation of the estimator of a parameter is equal to the parameter. By minimum variance, we mean that the estimator has the smallest variance among all unbiased estimators of the parameter (see [42]).

$t = 1, 2, \dots, n$, the log likelihood function at the sample is given by:

$$\mathcal{L}(\mathbf{X}_1, \dots, \mathbf{X}_n, \theta) = \sum_{t=1}^n \ln f(\mathbf{X}_t, \theta). \quad (4.3)$$

The maximum likelihood based estimators of θ are obtained by maximizing $\mathcal{L}(\cdot)$,

$$\hat{\theta}_{ml} = \operatorname{argmax}_{\theta} \mathcal{L}(\mathbf{X}_1, \dots, \mathbf{X}_n, \theta) = \operatorname{argmax}_{\theta} \sum_{t=1}^n \ln(f(\mathbf{X}_t, \theta)). \quad (4.4)$$

Now, for the models we adopt, the CCF, $\phi(s, \theta, \mathbf{X}_{t+1}|\mathbf{X}_t)$, of the sample is known, often in closed-form, as an exponential of an affine function of \mathbf{X}_t . Thus the conditional density function of \mathbf{X}_{t+1} given \mathbf{X}_t can be obtained by the Fourier transform of the CCF:

$$f(\mathbf{X}_{t+1}, \theta|\mathbf{X}_t) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} \phi(\mathbf{s}, \theta, \mathbf{X}_{t+1}|\mathbf{X}_t) e^{-i\mathbf{s} \cdot \mathbf{X}_{t+1}} d\mathbf{s}. \quad (4.5)$$

One can use the standard ML estimation based on this conditional density function to obtain ML-CCF estimators of the sample as:

$$\hat{\theta}_{CCF} = \operatorname{argmax}_{\theta} \sum_{t=1}^{n-1} \ln(f(\mathbf{X}_{t+1}, \theta|\mathbf{X}_t)). \quad (4.6)$$

Take Model 1a (3.6) as an example. The conditional density function of X_{t+1} given X_t of the sample is of the form:

$$\begin{aligned} f(X_{t+1}, \theta|X_t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(s, \theta, X_{t+1}|X_t) e^{-isX_{t+1}} ds \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isY_t} h(\theta, s) ds, \end{aligned} \quad (4.7)$$

where

$$Y_t = (X_{t+1} - \alpha) - e^{-\kappa}(X_t - \alpha), \quad (4.8)$$

and

$$\begin{aligned} h(\theta, s) &= \exp\left(-\frac{\sigma^2 s^2}{4\kappa}(1 - e^{-2\kappa}) + \frac{i\omega(1 - 2\psi)}{\kappa}(\arctan(\gamma s e^{-\kappa}) - \arctan(\gamma s))\right. \\ &\quad \left. + \frac{\omega}{2\kappa} \ln\left(\frac{1 + \gamma^2 s^2 e^{-2\kappa}}{1 + \gamma^2 s^2}\right)\right). \end{aligned} \quad (4.9)$$

To assist in computing this integral (4.7) we define

$$F(Y_t, \theta) := f(X_{t+1}, \theta | X_t) = \frac{1}{2\pi} \lim_{R \rightarrow \infty} \int_{-R}^R e^{-isY_t} h(\theta, s) ds. \quad (4.10)$$

Notice that $|h(\theta, s)|$ is continuous in s and $|h(\theta, s)| \leq -\frac{\sigma^2 s^2}{4\kappa} (1 - e^{-2\kappa})$. Thus one can truncate the integral to a finite interval $[-R, R]$ outside of which the function $h(\theta, s)$ to be integrated is negligibly small. Then, for this choice of R ,

$$F(Y_t, \theta) \approx \frac{1}{2\pi} \int_{-R}^R e^{-isY_t} h(\theta, s) ds. \quad (4.11)$$

Also, one can discretize Y_t into M subintervals such that:⁶

$$\begin{aligned} Y_n &= n\Delta Y_t = n\left(\frac{Y_t}{M}\right), \\ s_k &= k\Delta s = k\left(\frac{R}{M}\right), \\ F(Y_t, \theta) &\approx \frac{1}{2\pi} \frac{R}{M} \sum'_{n=-M}^{M-1} \left(e^{-ink \frac{RY_t}{M^2}} h\left(\theta, \frac{nR}{M}\right) \right). \end{aligned} \quad (4.12)$$

If we arrange $\frac{RY_t}{M} = 2\pi$, then we have

$$F(Y_t, \theta) \approx \frac{1}{2\pi} \frac{R}{M} \sum'_{n=-M}^{M-1} \left(e^{-ink \frac{2\pi}{M}} h\left(\theta, \frac{nR}{M}\right) \right). \quad (4.13)$$

One can approximate $F(Y_t, \theta)$ by the discrete Fourier transform (DFT) of $h(\theta, \frac{nR}{M})$, and the integral in Equation (4.7) can be estimated on a suitable grid of s values by a fast Fourier transform (FFT) algorithm.

4.3 ML-MCCF Estimators

ML-CCF estimators are asymptotically efficient if all of the state variables are observable. But for those multi-factor models with unobservable state variables such as

⁶Here we use the compound trapezoidal rule to approximate the integral, and $\sum'_i (A_i)$ denotes the sums of the A_i with the first and last term halved.

Model 3a and Model 3b, ML-CCF estimators cannot be obtained directly. If option prices are available, implied volatilities can be calculated from option prices observed in the market. Various numerical methods have been proposed for estimating implied volatility functions from option prices (see [45], [46], [47]). Then one can use those values as the data of volatilities and implement ML-CCF estimation.

But in our case, option prices are not available. Following Chacko and Viceira (1999) we can integrate the unobservable variable (volatility) from the joint CCF of the log price and the volatility, and set $s_v = 0$ to get the so-called marginal CCF. Based on this marginal CCF, we can implement a ML based estimation. This is somewhat similar to LML-CCF or Singleton's SMM method. But LML-CCF estimation not only keeps $s_v = 0$ but also needs to utilize the volatility information (not workable in our case). Singleton's SMM method integrates out the unobservable variable in the CCF by simulation. This requires a huge number of simulated paths of the volatility and can be quite time-consuming. Furthermore, this induces an estimation bias due to the discretization used in the simulation (see [13]). Meanwhile, compared to the SGMM method that we will introduce later on, ML-MCCF estimation avoids the so-called *ad hoc* moment conditions selection problem and is easier to implement in case of stochastic volatility models.

Take Model 3a as an example. Recall that the volatility follows a square-root process such as

$$dV_t = \kappa_v(\alpha_v - V_t)dt + \sigma_v\sqrt{V_t}dW_v. \quad (4.14)$$

The infinitesimal generator of the square-root process is:⁷

$$\mathcal{L}f(v) = \frac{\sigma_v^2 v}{2} \frac{\partial^2 f}{\partial v^2} + \kappa_v(\alpha_v - v) \frac{\partial f}{\partial v}. \quad (4.15)$$

Let μ_t be the distribution function of V_t , then it solves the forward Kolmogorov equation (4.16):⁸

$$\mu_t(\mathcal{L}f) = \frac{d}{dt} \mu_t(f), \quad (4.16)$$

with $\mu_t(f) := \int f(v) d\mu_t$.

In particular, let μ be the stationary characteristic function of the volatility. In this case, with $f(v) = e^{iuv}$ and $\hat{\mu}(u) := \mu(e^{iuv})$, we have

$$\begin{aligned} \mathcal{L}e^{iuv} &= -\frac{\sigma_v^2 v}{2} u^2 e^{iuv} + i\kappa_v(\alpha_v - v) u e^{iuv} \\ &= (iv(\frac{i\sigma_v^2 u^2}{2} - \kappa_v u) + i\kappa_v \alpha_v u) e^{iuv}, \end{aligned} \quad (4.17)$$

and

$$\begin{aligned} \mu(\mathcal{L}e^{iuv}) &= (\frac{i\sigma_v^2 u^2}{2} - \kappa_v u) \int i v e^{iuv} d\mu + i\kappa_v \alpha_v u \int e^{iuv} d\mu \\ &= (\frac{i\sigma_v^2 u^2}{2} - \kappa_v u) \frac{d\hat{\mu}(u)}{du} + i\kappa_v \alpha_v u \hat{\mu}(u). \end{aligned} \quad (4.18)$$

Because $\frac{d\mu(\cdot)}{dt} = 0$, we have

$$(\frac{i\sigma_v^2 u}{2} - \kappa_v) \frac{d\hat{\mu}(u)}{du} + i\kappa_v \alpha_v \hat{\mu}(u) = 0, \quad (4.19)$$

with $\hat{\mu}(0) = 1$. Then the solution for (4.19) has the form:

$$\hat{\mu}(u) = (1 - \frac{i u \sigma_v^2}{2 \kappa_v})^{-2 \kappa_v \alpha_v / \sigma_v^2}. \quad (4.20)$$

Recall that the joint CCF of the log price and volatility in this model is defined as

$$\begin{aligned} &\phi(s_x, s_v, \theta, X_T, V_T | X_t, V_t) \\ &= \exp(A(s_x, s_v, t, T, \theta) X_t + B(s_x, s_v, t, T, \theta) V_t + C(s_x, s_v, t, T, \theta)), \end{aligned} \quad (4.21)$$

⁷Taken from Rama Cont [48].

⁸For more details about forward Kolmogorov equation, see Paul Wilmott [19].

where $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ are the solutions of system (Equation (3.36)). As the stochastic volatility V_t is unobservable, we cannot estimate the parameters of stochastic models directly from the joint CCF of the log price and volatility. Let us define the marginal CCF as

$$\begin{aligned}
& \phi(s_x, \theta, X_T | X_t) \\
&= \int_0^\infty \phi(s_x, 0, \theta, X_T, V_T | X_t, V_t) d\mu \\
&= e^{A(s_x, 0, t, T, \theta)X_t + C(s_x, 0, t, T, \theta)} \int_0^\infty e^{B(s_x, 0, t, T, \theta)V_t} d\mu \\
&= e^{A(s_x, 0, t, T, \theta)X_t + C(s_x, 0, t, T, \theta)} \hat{\mu}(-iB(s_x, 0, t, T, \theta)).
\end{aligned} \tag{4.22}$$

Applying Equation (4.20), we obtain the marginal CCF of the form

$$\phi(s_x, \theta, X_T | X_t) = e^{A(s_x, 0, t, T, \theta)X_t + C(s_x, 0, t, T, \theta)} \left(1 - \frac{B(s_x, 0, t, T, \theta)\sigma_v^2}{2\kappa_v}\right)^{-2\kappa_v\alpha_v/\sigma_v^2}. \tag{4.23}$$

Through the Fourier transform, the marginal conditional density function is given by

$$f(X_{t+1}, \theta | X_t) = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(s_x, \theta, X_{t+1} | X_t) e^{-is_x X_{t+1}} ds. \tag{4.24}$$

Then, given a sample $\{X_t, t = 1, \dots, n\}$, one can implement the maximum likelihood estimation based on this marginal distribution of the observed variables (electricity prices), and obtain ML-MCCF estimators as

$$\hat{\theta}_{MCCF} = \operatorname{argmax}_{\theta} \sum_{t=1}^{n-1} \ln(f(X_{t+1}, \theta | X_t)). \tag{4.25}$$

Notice that since we only rely on the level of the electricity prices in the previous period, we lose efficiency. And the point estimations (including the SGMM estimators we will discuss later on), as pointed out by Chacko and Viceira (1999), are biased and inconsistent (see [13]). What's more, the theoretical value for the

bias is hard to calculate as we don't have closed forms for $B(\cdot)$ and $C(\cdot)$. Following Chacko and Viceira (1999), we try to correct the bias by a bootstrap method. Specifically, we simulate 500 paths with a given parameter θ_0 . For each path, there are 19,704 hourly observations (same length as the actual data). The estimates $\hat{\theta}_i$, $i = 1, \dots, n$, obtained from the simulated paths, result in a distribution for each parameter. We will regard the difference between the mean of those estimates and the given parameter as the bias, i.e.,

$$\text{bias} = \theta_0 - \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i, \quad (4.26)$$

with $n = 500$ in our setting. We will try adjusting the estimates from the actual data by this bias in the next chapter.

4.4 Spectral GMM Estimators

In this section, we will describe the SGMM estimators constructed by Chacko and Viceira (1999). This method is essentially GMM in a complex setting. GMM estimation is one of the most fundamental estimation methods in statistics and econometrics, especially after Hansen's influential paper [49] appeared in 1982. Unlike ML estimation which requires the complete specification of the model and its probability distribution, full knowledge of the specification and strong distributional assumptions are not required for GMM estimation. GMM estimators are best suited to study models that are only partially specified, and they are attractive alternatives to likelihood-type estimators.

We start with some basics of GMM, and define the moment conditions and mo-

ment functions first.⁹

Definition 4.1 *Suppose that we have a set of random variables $\{x_t, t = 1, 2, \dots\}$. Let $\theta = \{\theta_1, \dots, \theta_k\}$ be an unknown tuple with true value θ_0 to be estimated; θ_0, θ in some parameter space Θ . Then the q -dimensional vector of functions $\mathbf{m}(x_t, \theta)$ is called an (unconditional) **moment function** if the following **moment conditions** hold:*

$$E[\mathbf{m}(x_t, \theta_0)] = \mathbf{0}. \quad (4.27)$$

Notice that θ is a k -tuple vector and $E[\mathbf{m}(x_t, \theta_0)] = \mathbf{0}$ consists of q equations. If one has as many moment conditions as parameters to be estimated ($q = k$), one can simply solve the k equations in k unknowns to obtain the estimates. If we have fewer moment conditions than unknowns ($q < k$), then we cannot identify θ . In this case, we can “create” more moment conditions by the so-called weighting functions (often termed “instruments” in the GMM literature, see [50]). If we have more functions than unknowns ($q > k$), then this is an over-identified problem. Such cases of over-identification can easily arise and the moment estimator is not well-defined. Different choices of moment conditions may lead to different estimates. GMM is a method to solve this kind of over-identification problem.

Let us take the standard linear regression model as an example and consider

$$y = \mathbf{x}'\theta_0 + \varepsilon. \quad (4.28)$$

Here y is the response variable, $\mathbf{x} = [x_1, x_2, \dots, x_k]'$ is a k -dimensional vector of regressors, x' (as before) is its transpose, and $\theta = [\theta_1, \dots, \theta_k]'$ is the unknown vector

⁹For more details see [50].

of parameters with true value θ_0 . We assume that ε has zero expectation and is uncorrelated with \mathbf{x} . Using the Law of Iterated Expectations¹⁰ we find that

$$E[\mathbf{x}\varepsilon] = E[E[\mathbf{x}\varepsilon|\mathbf{x}]] = E[\mathbf{x}E[\varepsilon|\mathbf{x}]] = \mathbf{0}.$$

Therefore, we can have the moment functions $\mathbf{m}((\mathbf{x}, y), \theta) = \mathbf{x}(y - \mathbf{x}'\theta)$. These moment functions are well defined, since, by the assumptions

$$E[\mathbf{m}((\mathbf{x}, y), \theta_0)] = E[\mathbf{x}(y - \mathbf{x}'\theta_0)] = E[\mathbf{x}\varepsilon] = \mathbf{0}. \quad (4.29)$$

Suppose $n > k$ observations on the response variable are available, say y_1, y_2, \dots, y_n . Along with each observed response y_t , we have a k -dimensional observation vector of regressors \mathbf{x}_t . We have exactly as many moment conditions as parameters to be estimated, since \mathbf{x}_t is a k -dimensional vector. If we assume that the strong law of large numbers holds then we have

$$\frac{1}{n} \sum_{t=1}^n \mathbf{m}((\mathbf{x}_t, y_t), \hat{\theta}_n) \rightarrow E[\mathbf{m}((\mathbf{x}, y), \theta_0)] = \mathbf{0}, \text{ almost surely.} \quad (4.30)$$

So the Method of Moments (MM) estimator for this model is just the solution of

$$\frac{1}{n} \sum_{t=1}^n \mathbf{x}_t(y_t - \mathbf{x}_t'\hat{\theta}_n) = \mathbf{0}, \quad (4.31)$$

which gives

$$\hat{\theta}_n = \left(\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^n \mathbf{x}_t y_t = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (4.32)$$

with $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ and $\mathbf{y} = [y_1, \dots, y_n]'$. Thus, the ordinary least squares (OLS) estimator is a MM estimator.

Notice that we specified relatively little information about the error term ε . For ML estimation we would be required to give the distribution of the error term ε , as well as the autocorrelation and *heteroskedasticity*,¹¹ which are also not required in

¹⁰For more details see [42].

¹¹Heteroskedasticity means that the variance of the errors is not constant across observations.

formulating the moment conditions.

Now instead of assuming that the error term has zero expectation on certain observed variables, we can specify the moment conditions directly by requiring the error term to be uncorrelated with certain observed “instruments”. Let’s consider the previous model again. This time we do not assume the error term has zero expectation, but that it is still uncorrelated to the regressors. Suppose we have a q -dimensional observed instrument \mathbf{z} , ($q \geq k$) and $E[\mathbf{z}\varepsilon] = \mathbf{0}$. Thus we have the moment conditions

$$E[\mathbf{z}\varepsilon] = E[\mathbf{z}(y - \mathbf{x}'\theta_0)] = \mathbf{0}, \quad (4.33)$$

and the moment functions

$$\mathbf{m}((\mathbf{x}, y, \mathbf{z}), \theta) = \mathbf{z}(y - \mathbf{x}'\theta). \quad (4.34)$$

If $q = k$, then this is also a well defined problem. let \mathbf{z}_t denote the corresponding k -dimensional observation vector of instrument to y_t . We assume that the strong law of large numbers holds so that we have

$$\frac{1}{n} \sum_{t=1}^n \mathbf{m}((\mathbf{x}_t, y_t, \mathbf{z}_t), \hat{\theta}_n) \rightarrow E[\mathbf{m}((\mathbf{x}, y, \mathbf{z}), \theta_0)] = \mathbf{0}, \text{ almost surely.} \quad (4.35)$$

Therefore we solve

$$\frac{1}{n} \sum_{t=1}^n \mathbf{z}_t(y_t - \mathbf{x}_t'\hat{\theta}_n) = \mathbf{0}, \quad (4.36)$$

which gives

$$\hat{\theta}_n = \left(\sum_{t=1}^n \mathbf{z}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^n \mathbf{z}_t y_t = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y}. \quad (4.37)$$

with $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$.

Definition 4.2¹² Suppose we have an observed sample $\{\mathbf{x}_t, t = 1, 2, \dots, n\}$ from a stochastic process \mathbf{x} . Assume that the moment conditions $E[\mathbf{m}(\mathbf{x}_t, \theta_0)] = \mathbf{0}$ hold. One can define the **critierion function** as

$$\mathbf{Q}_n(\theta) = \left[\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\mathbf{x}_t, \theta) \right]' \mathbf{W}_n \left[\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\mathbf{x}_t, \theta) \right], \quad (4.38)$$

where \mathbf{W}_n is a $q \times q$ symmetric positive definite matrix. The **GMM estimator** $\hat{\theta}_n$ of θ associated with \mathbf{W}_n is the solution to the problem:

$$\begin{aligned} \hat{\theta}_n &= \underset{\theta}{\operatorname{argmin}} \quad \mathbf{Q}_n(\theta) \\ &= \underset{\theta}{\operatorname{argmin}} \quad \left[\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\mathbf{x}_t, \theta) \right]' \mathbf{W}_n \left[\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\mathbf{x}_t, \theta) \right]. \end{aligned} \quad (4.39)$$

Consider the linear regression model with instruments again, and suppose that we have $q > k$ moment functions this time. Suppose we choose

$$\mathbf{W}_n = \left(\frac{1}{n} \sum_{t=1}^n \mathbf{z}_t \mathbf{z}_t' \right)^{-1} = n(\mathbf{Z}'\mathbf{Z})^{-1}. \quad (4.40)$$

Then the criterion function is

$$\mathbf{Q}_n(\theta) = n^{-1}(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\theta)'(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\theta). \quad (4.41)$$

Differentiating with respect to θ gives the first order conditions

$$\frac{\partial \mathbf{Q}_n(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_n} = n^{-1} 2\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\hat{\theta}_n) = \mathbf{0}, \quad (4.42)$$

and solving for $\hat{\theta}_n$ gives

$$\hat{\theta}_n = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \quad (4.43)$$

¹² Taken from [50].

This is the standard instrument variable estimator for the case where there are more instruments than regressors.

Now, the definition of the CCF of the sample implies that

$$E[\exp(is \cdot \mathbf{X}_T) - \phi(\mathbf{s}, \theta, \mathbf{X}_T | \mathbf{X}_t)] = \mathbf{0}, \quad \mathbf{s} \in \mathbb{R}^n. \quad (4.44)$$

By taking real and imaginary parts of this function, we get the following pair of moment conditions:

$$\begin{aligned} E[\operatorname{Re}(\exp(is \cdot \mathbf{X}_T) - \phi(\mathbf{s}, \theta, \mathbf{X}_T | \mathbf{X}_t))] &= \mathbf{0}, \\ E[\operatorname{Im}(\exp(is \cdot \mathbf{X}_T) - \phi(\mathbf{s}, \theta, \mathbf{X}_T | \mathbf{X}_t))] &= \mathbf{0}. \end{aligned} \quad (4.45)$$

Thus one can define a set of moment functions $m(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t)$ as follows:¹³

$$\begin{aligned} \mathbf{m}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) &= \boldsymbol{\varepsilon}_t(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) = \begin{bmatrix} \boldsymbol{\varepsilon}_t^{\operatorname{Re}}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) \\ \boldsymbol{\varepsilon}_t^{\operatorname{Im}}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) \end{bmatrix}, \\ \boldsymbol{\varepsilon}_t^{\operatorname{Re}}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) &:= \operatorname{Re}(\boldsymbol{\varepsilon}_t(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t)) = \operatorname{Re}(\exp(is \cdot \mathbf{X}_T) - \phi(\mathbf{s}, \theta, \mathbf{X}_T | \mathbf{X}_t)), \\ \boldsymbol{\varepsilon}_t^{\operatorname{Im}}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) &:= \operatorname{Im}(\boldsymbol{\varepsilon}_t(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t)) = \operatorname{Im}(\exp(is \cdot \mathbf{X}_T) - \phi(\mathbf{s}, \theta, \mathbf{X}_T | \mathbf{X}_t)). \end{aligned} \quad (4.46)$$

¹³ $\operatorname{Re}(A)$ denotes the real part of A , while $\operatorname{Im}(A)$ denotes the imaginary part of A .

More generally, we can add a set of “instruments” or “weighting functions” to obtain more moment restrictions. So one can define the moment function based on the CCF as¹⁴

$$\mathbf{m}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) = \boldsymbol{\varepsilon}_t(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) \otimes \mathbf{p}(\mathbf{X}_t), \quad (4.47)$$

where $\mathbf{p}(\mathbf{X}_t)$ are “instruments” independent of $\boldsymbol{\varepsilon}_t(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t)$. The SGMM estimator is of the form:

$$\hat{\theta}_{SGMM} = \underset{\theta}{\operatorname{argmin}} \left[\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) \right]' W_n \left[\frac{1}{n} \sum_{t=1}^n \mathbf{m}(\mathbf{s}, \theta, \mathbf{X}_T, \mathbf{X}_t) \right], \quad \theta \in \Theta. \quad (4.48)$$

Just as for other GMM estimators, the asymptotic variance of the SGMM estimator is minimized with the optimal weighting matrix $W_n = S^{-1}$, where S is the covariance matrix of the moment functions (see [13]). Under the usual regularity conditions,¹⁵ according to Chacko and Viceira (1999), the SGMM estimator, $\hat{\theta}_{SGMM}$, inherits the optimality properties of GMM estimators such as consistency and asymptotic normality (see [49], [13]).

We now can apply SGMM to estimate stochastic volatility models like Model 3a

¹⁴Let \mathbf{A} , \mathbf{B} be $K \times L$, $M \times N$ matrices with elements indexed as

$$\begin{aligned} A[k, l], \quad k = 0, 1, \dots, K-1, \quad l = 0, 1, \dots, L-1, \\ B[m, n], \quad m = 0, 1, \dots, M-1, \quad l = 0, 1, \dots, N-1. \end{aligned}$$

We define the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ to be a $KM \times LN$ matrix

$$\mathbf{A} \otimes \mathbf{B} := \begin{bmatrix} A[0, 0]\mathbf{B} & A[0, 1]\mathbf{B} & \cdots & A[0, L-1]\mathbf{B} \\ A[1, 0]\mathbf{B} & A[1, 1]\mathbf{B} & \cdots & A[1, L-1]\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A[K-1, 0]\mathbf{B} & A[K-1, 1]\mathbf{B} & \cdots & A[K-1, L-1]\mathbf{B} \end{bmatrix}$$

with elements

$$(A \otimes B)[m + kM, n + lN] := A[k, l]B[m, n].$$

¹⁵We assume that Hansen’s regularity conditions are satisfied. For more details see [49].

and Model 3b. Recall that the marginal CCF of the sample is given by:

$$\phi(s_x, \theta, X_{t+1}|X_t) = e^{A(s_x, 0, t, \theta)X_t + C(s_x, 0, t, \theta)} \left(1 - \frac{B(s_x, 0, t, \theta)\sigma_v^2}{2k_v}\right)^{-2\kappa_v \alpha_v / \sigma_v^2}, \quad (4.49)$$

where $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ are the solutions of system (3.36) and (3.43) for Model 3a and Model 3b respectively. Given a sample $\{X_t, t = 1, \dots, T\}$, we have moment functions as follows:

$$\mathbf{m}(s_x, X_t, \theta) = \boldsymbol{\varepsilon}_t(s_x, X_t, \theta) = \begin{bmatrix} \boldsymbol{\varepsilon}_t^{\text{Re}}(s_x, X_t, \theta) \\ \boldsymbol{\varepsilon}_t^{\text{Im}}(s_x, X_t, \theta) \end{bmatrix}, \quad (4.50)$$

$$\boldsymbol{\varepsilon}_t^{\text{Re}}(s_x, X_t, \theta) = (\cos(s_x X_{t+1}) - \text{Re}(\phi(s_x, \theta, X_{t+1}|X_t))) \otimes \mathbf{p}(X_t),$$

$$\boldsymbol{\varepsilon}_t^{\text{Im}}(s_x, X_t, \theta) = (\sin(s_x X_{t+1}) - \text{Im}(\phi(s_x, \theta, X_{t+1}|X_t))) \otimes \mathbf{p}(X_t).$$

with $\phi(s_x, \theta, X_{t+1}|X_t)$ defined as Equation (4.49). Following Chacko and Viceira (1999), one can compute the n -th conditional moment by simple substitution of $s_x = n$ into Equation (4.50).¹⁶

In the next chapter, in order to construct the moment functions, we use the first six spectral moments by setting $s_x = 1, 2, \dots, 6$ and $\mathbf{p}(X_t)$ as a T -dimensional vector of 1s. Again, as we only explored the information in the marginal CCF, the estimates we obtain are biased and inconsistent (see [50]).¹⁷

¹⁶Paraphrasing [50], the selection of an appropriate set of s_x is crucial for the efficiency of GMM estimation. A poor selection of s_x which results in a poor choice of moment conditions may lead to very inefficient estimators and can even cause identification problems. This is known as the *ad hoc* choice of moment conditions problem in the GMM literature.

According to Carrasco *et al.* (see [29]), when s_x goes through all real numbers, i.e., when the number of moment conditions increases to infinity, the conditional density can be recovered. The corresponding GMM estimator based on a continuum of moments conditions yields ML efficiency. But the size of the weighting matrix is $n \times n$ where n is the sample size. Therefore the computational burden increases dramatically as the sample size increases. Also, the estimators will deteriorate for large sample sizes as the numerical errors associated with large matrix operations, according to Mikhail Chernov (one of the authors of [29]).

¹⁷We will use the GMM and MINZ Program Libraries for Matlab, by Michael T. Cliff, to perform the estimation.

Chapter 5

Model Comparison

This chapter gives some empirical comparisons among the models we introduced in Chapter 3. Three data sets—the hourly electricity prices (Hourly EP), on-peak daily average electricity prices (Peak EP), and off-peak daily average electricity prices (Offpeak EP)—from January 1, 2002 to March 31, 2004, are used to evaluate the price models.¹

5.1 Data Description

A summary of statistics for the hourly electricity price series from January 1, 2002 to March 31, 2004 is presented in Table 5.1. The statistics reported here are for electricity prices (P), the change in electricity prices (dP), the logarithm of electricity prices ($\ln(P)$), and the log returns of electricity prices ($d\ln(P)$), from one hour to the next. Clearly, price series are skewed with excess kurtosis.

The descriptive statistics for the on-peak/off-peak daily average electricity prices during the period January 1, 2002 to March 31, 2004 are shown in Tables 5.2 and 5.3 respectively. The data is daily in frequency. As might be expected, Hourly EP has more volatility, larger skewness and higher kurtosis than the daily average

¹The hourly electricity prices were obtained from the public website of the Alberta Electric System Operator. Altogether, we have 19,704 observations at an hourly frequency. The on-peak and off-peak daily average electricity price were computed according to the Alberta Energy and Utilities Board. Thus we have 569 observations for the on-peak period, and 821 observations for the off-peak period.

| | mean | Std.Dev | Skewness | Kurtosis | Minimum | Maximum |
|------------|---------|---------|----------|----------|-----------|----------|
| P | 58.5911 | 64.2576 | 5.1325 | 46.6263 | 0.0100 | 999.9900 |
| dP | -0.0042 | 54.4378 | -0.4207 | 53.1331 | -862.1000 | 826.3900 |
| $\ln(P)$ | 3.5922 | 0.8359 | 0.1089 | 4.1445 | -4.6052 | 6.9077 |
| $d \ln(P)$ | -0.0000 | 0.5529 | -0.0133 | 8.4214 | -6.0426 | 6.0234 |

Table 5.1: Descriptive statistics of Hourly EP

Note: here and in the remainder of this thesis, the column labeled Std.Dev reports the standard deviation. Skewness and Kurtosis are the third, fourth moment around the mean, namely skewness = $\frac{E[X-E[X]]^3}{[var[X]]^{1.5}}$, kurtosis = $\frac{E[X-E[X]]^4}{[var[X]]^2}$. For a normal distribution, skewness is equal to 0 and kurtosis is equal to 3.

| | mean | Std.Dev | Skewness | Kurtosis | Minimum | Maximum |
|------------|---------|---------|----------|----------|-----------|----------|
| P | 68.3287 | 43.0324 | 2.8119 | 16.6829 | 12.2264 | 413.9657 |
| dP | -0.0150 | 46.5599 | 0.0897 | 11.1163 | -283.8329 | 255.9136 |
| $\ln(P)$ | 4.0763 | 0.5340 | 0.1347 | 3.6559 | 2.5036 | 6.0258 |
| $d \ln(P)$ | -0.0004 | 0.5225 | 0.0910 | 5.0479 | -2.3969 | 2.1600 |

Table 5.2: Descriptive statistics of Peak EP

on-peak/off-peak data. And the descriptive statistics for Peak EP are larger than Offpeak EP.

Figure 5.1 plots the histogram of the log returns of Hourly EP over the period January 1, 2002 to March 31, 2004. There is a big spike in the middle which is a quantization effect (mainly due to rounding to $\pm .01$). This is hard to be captured by most models. In order to get reasonable estimates, we remove systematic day to day variations from data prior to fitting. In this way, we can reduce the influence of hourly price patterns and compensate for quantization effects. Specifically, if P_t is the actual data at the i -th hour, $i = 1, \dots, 24$, then the “deseasonalized” data X_t is

| | mean | Std.Dev | Skewness | Kurtosis | Minimum | Maximum |
|------------|---------|---------|----------|----------|-----------|----------|
| P | 39.6513 | 29.8016 | 2.6022 | 14.9079 | 6.5713 | 299.2710 |
| dP | -0.0116 | 25.5697 | -0.2651 | 12.7988 | -200.9500 | 128.0813 |
| $\ln(P)$ | 3.4647 | 0.6445 | 0.2188 | 2.7684 | 1.8827 | 5.7013 |
| $d \ln(P)$ | -0.0005 | 0.4710 | 0.1014 | 4.9058 | -2.1208 | 2.2547 |

Table 5.3: Descriptive statistics of Offpeak EP

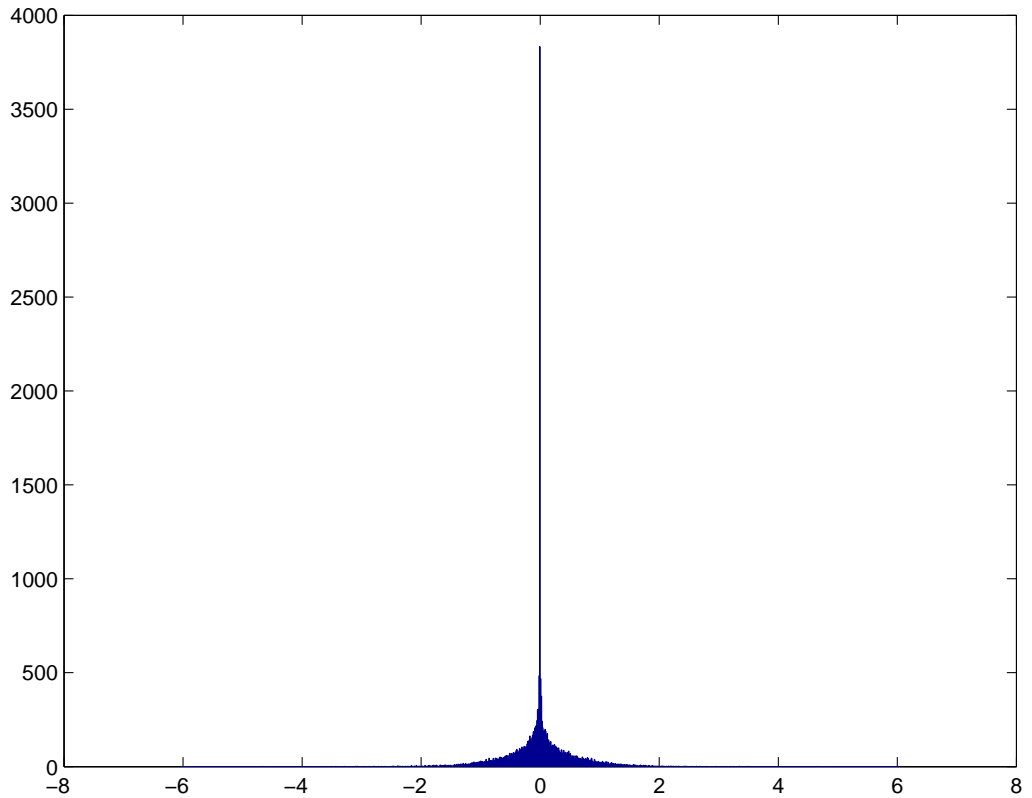


Figure 5.1: Histogram of the log returns of Hourly EP

Note: the price series we obtained from the public website of the Alberta Electric System Operator have been rounded to the nearest two decimals. A significant amount of log returns are inside the range -0.01 to 0.01.

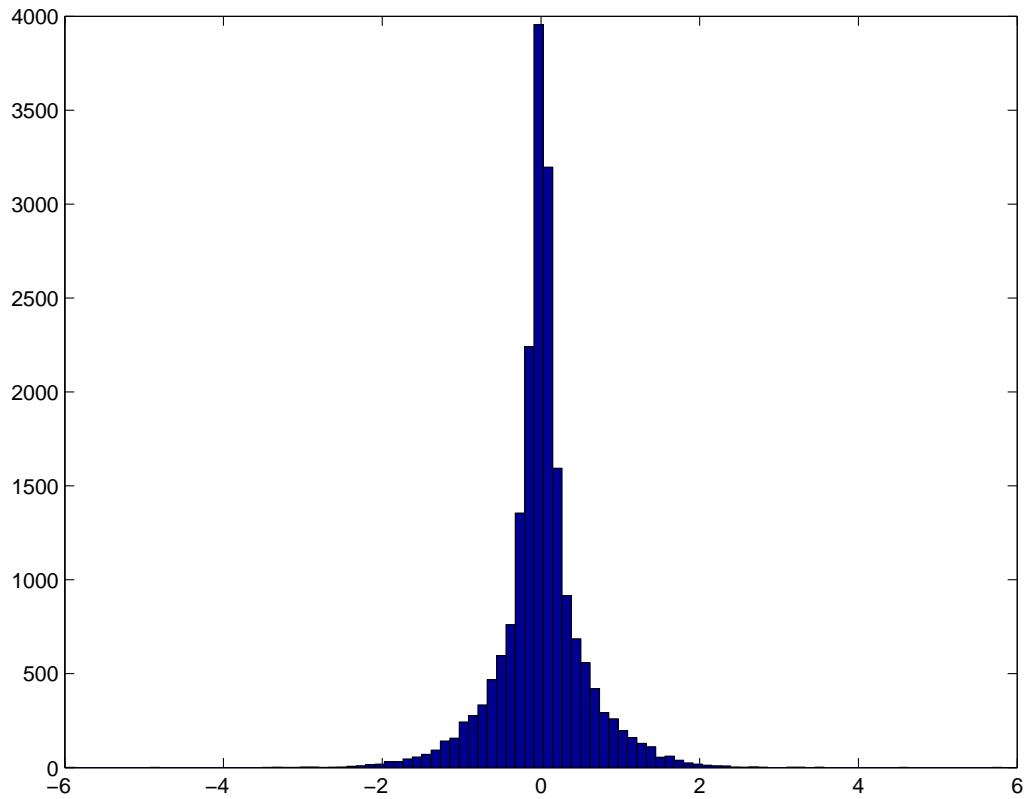


Figure 5.2: Histogram of the changes of the deseasonalized data (dX)

obtained by:

$$X_t = \ln(P_t) - \text{Mean of the log price at the } i\text{-th Hour.} \quad (5.1)$$

Figure 5.2 plots the histogram of the changes of the “deseasonalized” data. Table 5.4 presents descriptive statistics for the log of Hourly EP ($\ln(P)$), the log returns of Hourly EP, “deseasonalized” data (X) and the changes of the deseasonalized data (dX).

| | mean | Std.Dev | Skewness | Kurtosis | Minimum | Maximum |
|------------|---------|---------|----------|----------|---------|---------|
| $\ln(P)$ | 3.5922 | 0.8359 | 0.1089 | 4.1445 | -4.6052 | 6.9077 |
| $d \ln(P)$ | -0.0000 | 0.5529 | -0.0133 | 8.4214 | -6.0426 | 6.0234 |
| X | -0.0000 | 0.7344 | 0.3104 | 4.9780 | -7.7314 | 3.5917 |
| dX | -0.0000 | 0.5362 | -0.0307 | 8.8230 | -5.9823 | 5.8074 |

Table 5.4: Descriptive statistics of deseasonalized Hourly EP

5.2 Calibration

This section presents estimates from fitting each of the six models to Hourly EP using the methodologies described in Chapter 4. In the remainder of this thesis, we will refer to Model 1a, Model 2a and Model 3a as the one-jump version models and refer to Model 1b, Model 2b and Model 3b as the two-jump version models. In all cases, Hourly EP were deseasonalized prior to fitting.²

For Model 1a, Model 2a, Model 1b and Model 2b, we perform calibration via ML-CCF estimation. As we have computed the CCF of X_{t+1} given X_t for each model in Chapter 3, just as for the example in Section 4.2, the conditional density of X_{t+1} given X_t can be recovered through the Fourier transform.³ The estimates are obtained by using standard ML estimation on these conditional densities. We used optimization toolbox in Matlab for the optimization. The optimizations converged in less than 20 steps for Hourly EP. Actual computer time required for each optimization is around 1 minute on a 2.46 GHz Pentium 4 PC. The resulting parameter estimates, the corresponding standard error (labeled as Std.) and t-ratio⁴ for Hourly EP via ML-CCF estimation are provided in Tables 5.5–5.10.

²In the remainder of this thesis, we just denote the deseasonalized Hourly EP as Hourly EP.

³All the routines are realized in Matlab, and are available on request.

⁴T-ratio is defined as the ratio of the estimate, to the estimate of standard deviation of the

Notice that the models are generally consistent with each other. The mean reverting rate κ is similar for all the models, and is close to 0.11. Thus the half-life of the mean-reverting process is about 2.7 hours.⁵ What's more, the jump parameters ω , ψ and γ for all the one-jump version models are quite similar. Also, the jump parameters ω_u , γ_u , ω_d and γ_d in the two-jump version models are similar. This is maybe because that the difference between the so obtained onpeak data and offpeak data is not significant in Alberta (refer to Figure [?]).

Meanwhile, we observe that the frequencies of positive jumps and negative jumps are quite close to each other. The probability of a positive jump ψ is roughly 50% in the one-jump version models. Likewise, the upward jump arrival rate ω_u is roughly equal to the downward jump arrival rate in the two-jump version models. According to Barz (1999), the derived balance between upward and downward jumps can be explained because we use the logarithm of the prices instead of using the spot prices. The magnitude of downward movements is amplified relative to upward movements. What's more, he suggested that care must be taken to ensure one does not use a misspecified model in this case (see [17]).

We also observe some different behaviour among these models. Incorporating the on-peak and off-peak effects in Model 2a does not change most of the estimates sampling distribution of the estimate. That is,

$$\text{t-ratio} = \frac{\text{Estimate}}{\text{Std.}}. \quad (5.2)$$

For large degrees of freedom (usually 30 or more), if the absolute value of the t-ratio is larger than 1.96, we can consider the parameter to be significant at the 95% level.

⁵Half-life is a key property of a mean-reverting process. It is the time that it takes for the price to revert half way back to its long-term mean level from its current level if no more random shocks arrive. Typically, for a mean-reverting process, we have the half-life is equal to $\ln(2)$ over the mean-reverting rate. For more details, see [23].

significantly compared with the estimates from Model 1a. Only the estimate for the long-term mean in Model 1b is much smaller than the estimate for the long-term mean in Model 1a. The on-peak coefficient α_1 is close to the off-peak coefficient α_2 in Model 2a. However, in Model 2b, this is not the case. Moreover, the probability of positive jumps is slightly more than the probability of negative jumps in Model 1a, while we have slightly more negative jumps than positive jumps in Model 1b.

The graphic outputs from the optimization process are reported in Figures 5.3–5.6. Each figure consists of three parts. The upper left plot is the empirical histogram of “deviations from the expected values”. For Model 1a and Model 1b, this is defined as:⁶

$$Y_t = (X_{t+1} - \alpha) - e^{-\kappa}(X_t - \alpha), \quad (5.3)$$

where X_t is the log price at time t , κ is the mean-reverting rate, and α is the long-term mean. It can be shown that the density function of Y_t is the same as $f(X_{t+1}, \theta|X_t)$ (with possibly a translated mean). Hence this histogram and the corresponding density should have the same shape. For Model 2a and Model 2b, it is defined as:

$$Y_t = (X_{t+1} - \alpha(t)) - e^{-\kappa}(X_t - \alpha(t)), \quad (5.4)$$

where

$$\alpha(t) = \alpha_1 \text{peak}_t + \alpha_2 \text{offpeak}_t, \quad (5.5)$$

with peak_t and offpeak_t defined in Equation (3.22). The upper right plot is the theoretical density of Y_t . For good estimates, the histograms of deviations from expected values should be similar to the theoretical conditional density plots. The lower plots are slices of the log-likelihood function versus each parameter. The

⁶Refer back to Equation (4.8).

asterisk on each curve is the estimate for the corresponding parameter. If the asterisk is at the peak of the curve, then the estimate is a good one. The number labeled below the x-axis is the width of each curve, which is the 95% confidence interval. All the histograms of deviations from expected values are very similar to the theoretical conditional density plots. Also, all of the asterisks (denotes the estimates) are at the peak of the curves.

For these stochastic volatility models (Model 3a and Model 3b), since the information of the volatility is not available, we cannot use ML-CCF estimation. However, we can explore the marginal CCF, and perform ML-MCCF estimation instead.⁷ It takes more than 20 minutes for the optimizations to converge on a 2.46 GHz Pentium 4 PC. The resulting parameter estimates, the corresponding standard error and t-ratio for Hourly EP via ML-MCCF estimation are provided in Tables 5.9 and 5.10. As pointed out by Chacko and Viceira (1999), these estimates are biased and inconsistent, and need to be modified. We tried to correct the bias by a bootstrap method as suggested by Chacko and Viceira (1999), and simulate 500 paths with a given parameter θ_0 .⁸ The given parameter, the mean of the estimates, mean bias and

⁷Refer back to Section 4.3.

⁸Unfortunately, we cannot obtain simulated paths with the same length as the actual data using the estimates in Tables 5.9 and 5.10 directly. Because $2\kappa_v\alpha_v$ is very close to σ_v^2 , it is very hard to guarantee the positiveness of the volatility process for a long path.

root mean square error (RMSE)⁹ for Model 3a and Model 3b are reported in Table 5.11 and Table 5.12 respectively. In comparison with the estimates, the RMSEs are pretty big, which indicates that the estimates are not reliable. We then decided not to use these biases to adjust our estimates.

We also report the graphic outputs from the optimization process in Figures 5.7–5.8. Similar to before, each figure consists of three parts. The upper left plot is the empirical histogram of deviations from the expected values. For Model 3a and Model 3b, deviation from the expected value is defined as:

$$Y_t = X_{t+1} - e^{-\kappa} X_t. \quad (5.6)$$

The upper right plot is the theoretical density of Y_t . The lower plots are slices of the log-likelihood function versus each parameter. The histograms of deviations from expected values are close to the theoretical conditional density plots. But not all of the asterisks are at the peak of the curves. The asterisk denoted for ρ is not at the peak. Also, the curve of the log-likelihood function versus the correlation coefficient (rho) in both models is quite flat (not convex). This indicates that this parameter is hard to estimate, and extra care should be taken into when computing this parameter. It seems that ρ tends to go towards -1, and this may result in higher log-likelihood. Therefore, we recomputed the log-likelihood function for both models

⁹Let $\hat{\theta}_i$ be the estimate for each sample path $i, i = 1, \dots, n$ and θ_0 be the true value, then mean bias is defined as:

$$\text{Mean Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta_0).$$

RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta_0)^2}.$$

| | Estimate | Std. | T-ratio |
|------------|----------|--------|---------|
| κ | 0.1124 | 0.0040 | 28.1000 |
| α | -0.2368 | 0.0250 | -9.4720 |
| σ^2 | 0.0144 | 0.0008 | 18.0000 |
| ω | 1.2260 | 0.0344 | 35.6395 |
| ψ | 0.5261 | 0.0061 | 86.2459 |
| γ | 0.3367 | 0.0058 | 58.0517 |

Table 5.5: Hourly EP parameters values for Model 1a

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 1a ($dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t + Q_t dP_t(\omega)$) to Hourly EP. All the estimates are hourly based. The standard errors are quite small. All the absolute values of the t-ratios are larger than 1.96, which indicates that all the parameters are likely significant.

fixing $\rho = -1$. The resulting parameter estimates, the corresponding standard error and t-ratio for Hourly EP via ML-MCCF estimation are provided in Tables 5.13 and 5.14. The log-likelihood values are really higher with $\rho = -1$ which verifies our hypothesis. As the spot price and the volatility are so strongly correlated, Model 3a and Model 3b may be misspecified models. Therefore, in the following sections, we will only consider the other four models.

We also carried out estimation for Model 3a and Model 3b, via SGMM estimation, by fitting Hourly EP. Unfortunately, the results that were obtained in the available time constraint were not very accurate, and are not presented here. In comparison with the other four models, the optimization processes of Model 3a and Model 3b are quite computationally intensive. This is because we have to solve the Riccati equations numerically during each iteration.

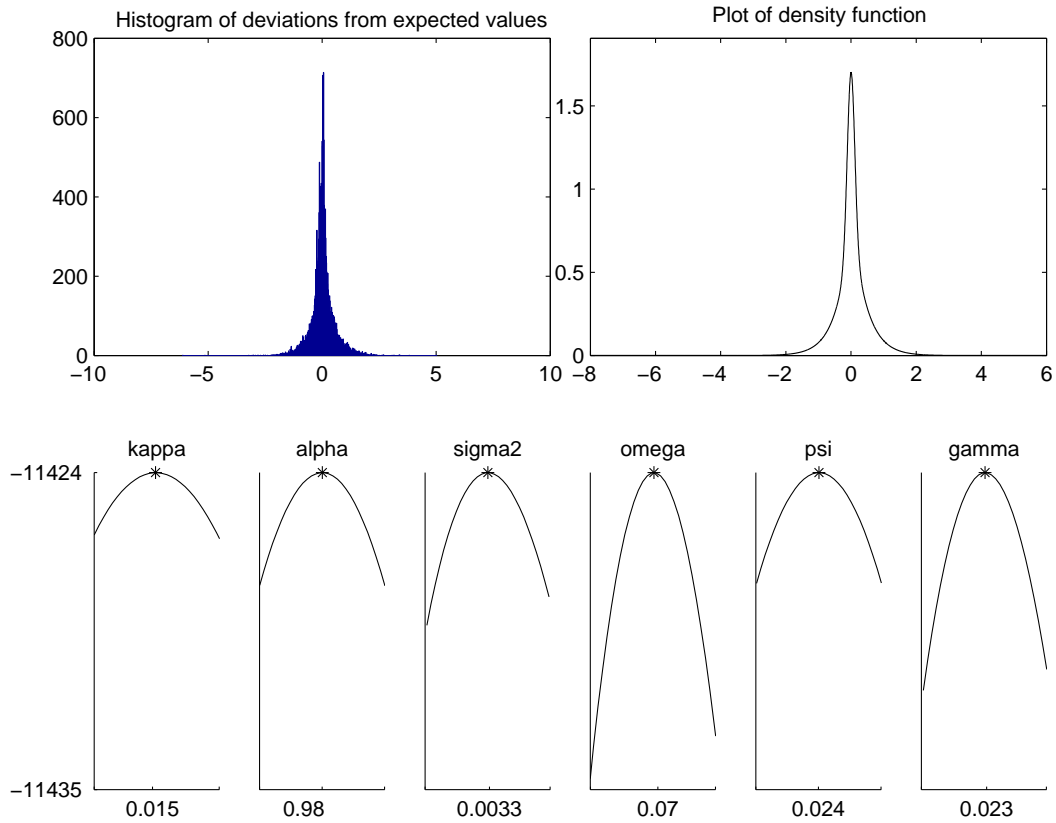


Figure 5.3: Results from optimization (Model 1a)

Note: these are the results of fitting Model 1a ($dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t + Q_t dP_t(\omega)$) to Hourly EP. If we get good estimates, the histogram of deviation of expected values (defined in (5.3)) should be very similar to the plot of the theoretical density function. From these plots, one can tell that Model 1a is a good fit to Hourly EP. Moreover, all the estimates are at the peak of each curve, which indicates that we have found the optimum points. These results are also consistent with the small standard errors in Table 5.5.

| | Estimate | Std. | T-ratio |
|------------|----------|--------|---------|
| κ | 0.1051 | 0.0041 | 25.6341 |
| α_1 | -0.2765 | 0.0278 | -9.9460 |
| σ^2 | 0.0136 | 0.0008 | 17 |
| ω | 1.2188 | 0.0328 | 37.1585 |
| ψ | 0.5228 | 0.0061 | 81.7049 |
| γ | 0.3384 | 0.0057 | 59.3684 |
| α_2 | -0.2009 | 0.0274 | -7.3321 |

Table 5.6: Hourly EP parameters values for Model 2a

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 2a ($dX_t = \kappa(\alpha(t) - X_t)dt + \sigma dW_t + Q_t dP_t(\omega)$) to Hourly EP. All the estimates are hourly based. The standard errors are quite small. All the absolute values of the t-ratios are larger than 1.96, which indicates that all the parameters are significant. In comparison with the estimates from Model 1a, incorporating the on-peak and off-peak effects does not change our estimates significantly. Moreover, the on-peak coefficient α_1 is close to the off-peak coefficient α_2 .

| | Estimate | Std. | T-ratio |
|------------|----------|--------|----------|
| κ | 0.1162 | 0.0042 | 27.6667 |
| α | -0.1405 | 0.0320 | -4.3906 |
| σ^2 | 0.0131 | 0.0009 | 14.5556 |
| ω_u | 0.6190 | 0.0187 | 33.1016 |
| γ_u | 0.3536 | 0.0073 | 48.4384 |
| ω_d | 0.6622 | 0.0299 | 22.1472 |
| γ_d | -0.3059 | 0.0084 | -36.4167 |

Table 5.7: Hourly EP parameters values for Model 1b

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 1b ($dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t + Q_t^u dP_t^u(\omega_u) + Q_t^d dP_t^d(\omega_d)$) to Hourly EP. All the estimates are hourly based. The standard errors are small. All the absolute values of the t-ratios are larger than 1.96, which indicates that all the parameters are significant. Notice that we have slightly more negative jumps than positive jumps, while, in Model 1a, the probability of positive jumps is slightly more than the probability of negative jumps. Furthermore, the estimate for the long-term mean in Model 1b is much smaller than the estimate for the long-term mean in Model 1a.

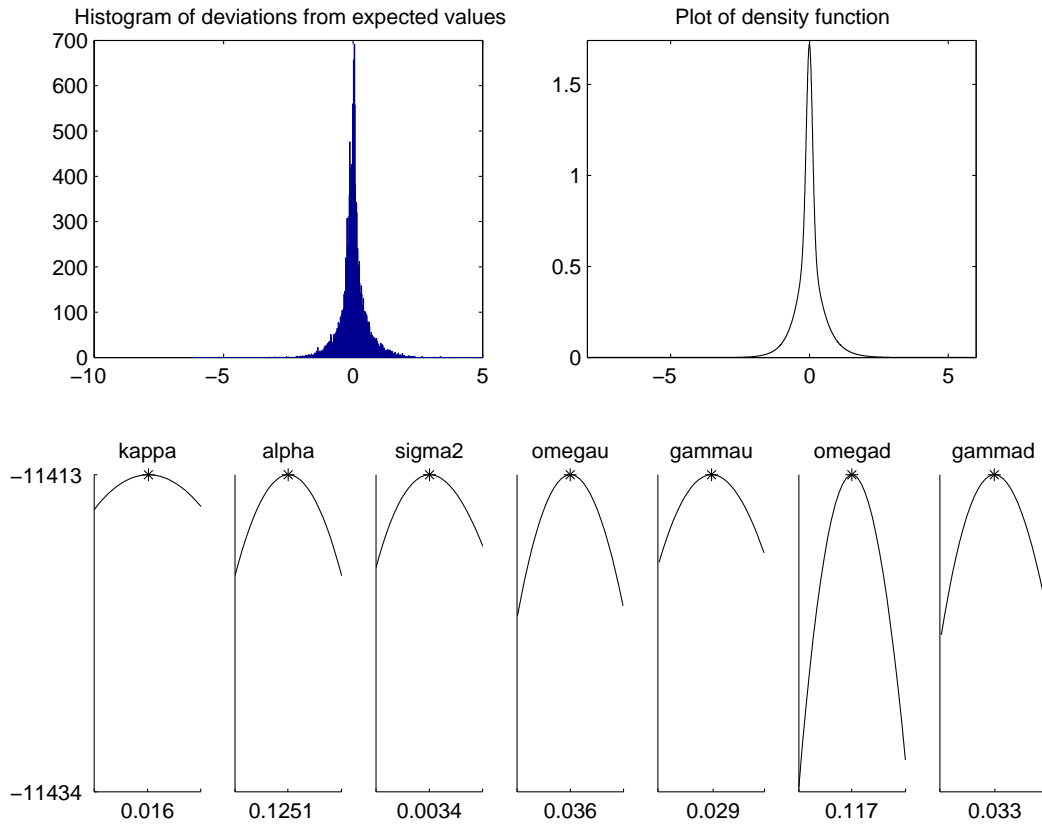


Figure 5.4: Results from optimization (Model 2a)

Note: these are the results of fitting Model 2a ($dX_t = \kappa(\alpha(t) - X_t)dt + \sigma dW_t + Q_t dP_t(\omega)$) to Hourly EP. If we get good estimates, the histogram of deviation of expected values (defined in (5.4)) should be very similar to the plot of the theoretical density function. From these plots, one can tell that Model 2a is a good fit to Hourly EP. Moreover, all the estimates are at the peak of each curve, which indicates that we have found the optimum points. These results are also consistent with the small standard errors in Table 5.6.

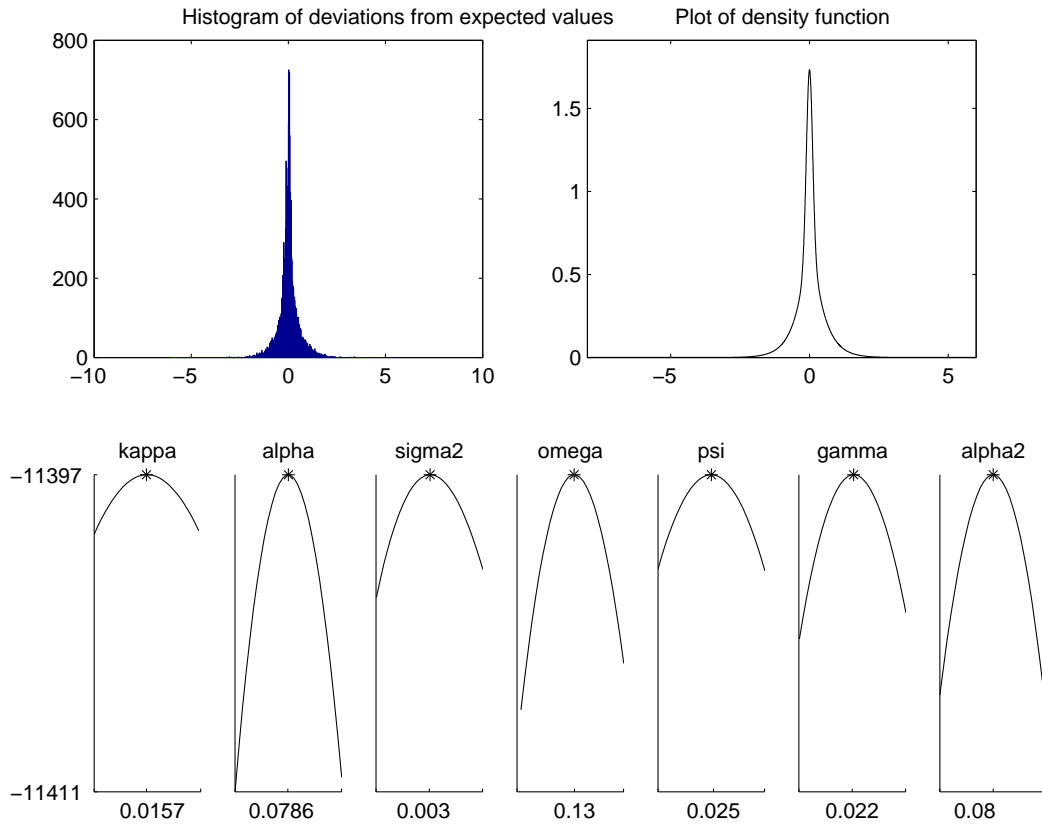


Figure 5.5: Results from optimization (Model 1b)

Note: these are the results of fitting Model 1b ($dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t + Q_t^u dP_t^u(\omega_u) + Q_t^d dP_t^d(\omega_d)$) to Hourly EP. If we get good estimates, the histogram of deviation of expected values (defined in (5.3)) should be very similar to the plot of the theoretical density function. From these plots, one can tell that Model 1b is a good fit to Hourly EP. Moreover, all the estimates are at the peak of each curve, which indicates that we have found the optimum points. These results are also consistent with the small standard errors in Table 5.7.

| | Estimate | Std. | T-ratio |
|------------|----------|--------|---------|
| κ | 0.1084 | 0.0041 | 26.1427 |
| α_1 | -0.1665 | 0.0338 | -4.9205 |
| σ^2 | 0.0121 | 0.0008 | 14.8066 |
| ω_u | 0.6096 | 0.0175 | 34.9320 |
| γ_u | 0.3569 | 0.0072 | 49.5694 |
| ω_d | 0.6725 | 0.0294 | 22.8429 |
| γ_d | -0.3043 | 0.0081 | 37.4514 |
| α_2 | -0.0870 | 0.0343 | -2.5346 |

Table 5.8: Hourly EP parameters values for Model 2b

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 2b ($dX_t = \kappa(\alpha(t) - X_t)dt + \sigma dW_t + Q_t^u dP_t(\omega_u) + Q_t^d dP_t(\omega_d)$) to Hourly EP. All the estimates are hourly based. The standard errors are quite small. All the absolute values of the t-ratios are larger than 1.96, which indicates that all the parameters are significant. In comparison with the estimates from Model 1b, incorporating the on-peak and off-peak effects does not change our estimates significantly. The on-peak coefficient α_1 is not that close to the off-peak coefficient α_2 as in Model 2a.

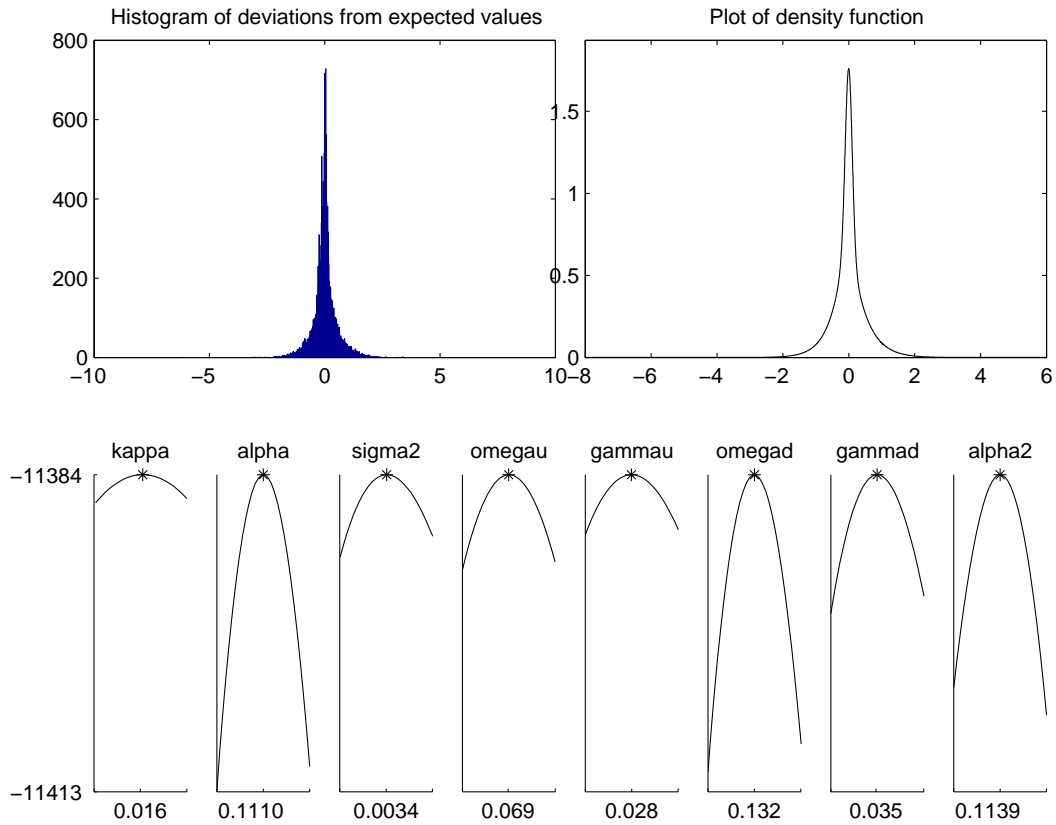


Figure 5.6: Results from optimization (Model 2b)

Note: these are the results of fitting Model 2b ($dX_t = \kappa(\alpha(t) - X_t)dt + \sigma dW_t + Q_t^u dP_t(\omega_u) + Q_t^d dP_t(\omega_d)$) to Hourly EP. If we get good estimates, the histogram of deviation of expected values (defined in (5.4)) should be very similar to the plot of the theoretical density function. From these plots, one can tell that Model 2b is a good fit to Hourly EP. Moreover, all the estimates are at the peak of each curve, which indicates that we have found the optimum points. These results are also consistent with the small standard errors in Table 5.8.

| | Estimate | Std. | T-ratio |
|----------------|-------------|--------|-----------|
| κ | 0.1152 | 0.0023 | 50.0870 |
| α | -0.5086 | 0.0413 | -12.3148 |
| ω | 1.0814 | 0.0170 | 63.6118 |
| ψ | 0.5781 | 0.0090 | 64.2333 |
| γ | 0.3522 | 0.0065 | 54.1846 |
| ρ | -0.9710 | 0.0014 | -693.5714 |
| κ_v | 0.5483 | 0.0315 | 17.4063 |
| α_v | 0.0261 | 0.0020 | 13.0500 |
| σ_v | 0.1529 | 0.0171 | 8.9415 |
| Log-likelihood | -11345.8013 | | |

Table 5.9: Hourly EP parameters values for Model 3a

Note: this table reports the estimates, the standard error of estimation, t-statistics and adjusted estimates (labeled as A.Est.) from fitting Model 3a ($d \begin{bmatrix} X_t \\ V_t \end{bmatrix} = \begin{bmatrix} \kappa(\alpha - X_t) \\ \kappa_v(\alpha_v - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{1 - \rho^2} \sqrt{V_t} & \rho \sqrt{V_t} \\ 0 & \sigma_v \sqrt{V_t} \end{bmatrix} \begin{bmatrix} dW_t \\ dW_v \end{bmatrix} + \begin{bmatrix} Q_t dP_t(\omega) \\ 0 \end{bmatrix}$) to Hourly EP. All the estimates are hourly based. The standard errors are still reasonably small. All the absolute values of the t-ratios are larger than 1.96, which indicates that all the parameters are significant.

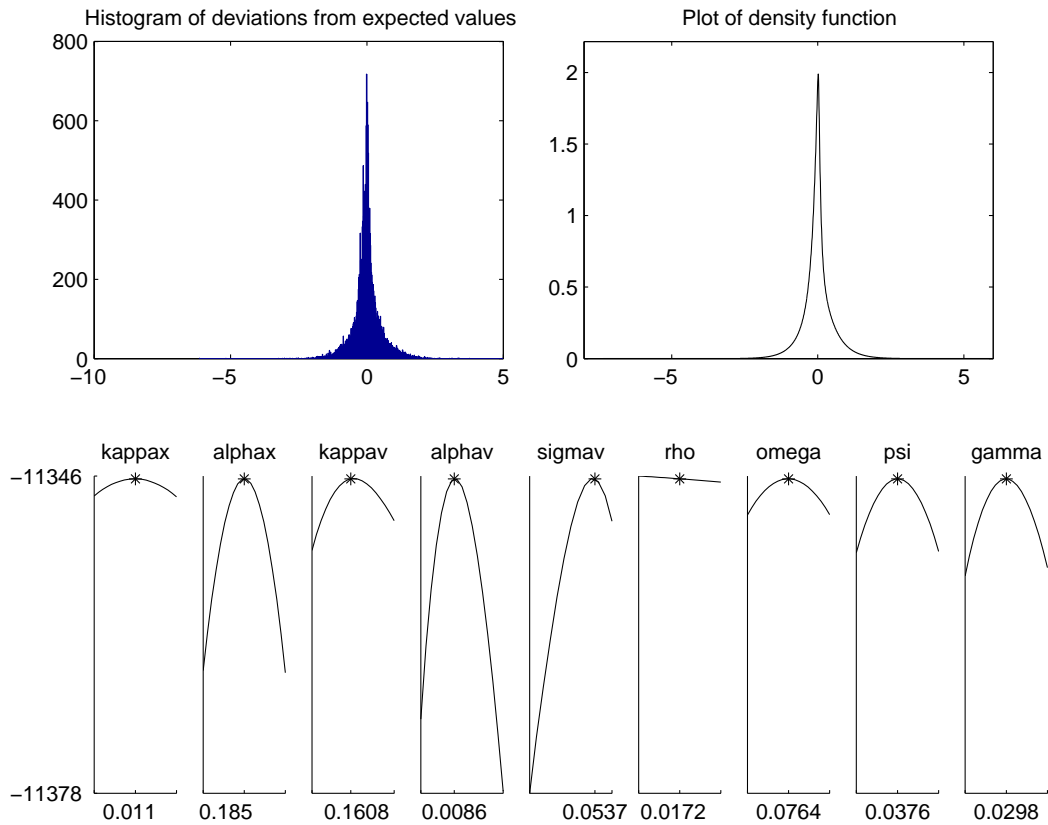


Figure 5.7: Results from optimization (Model 3a)

Note: these are the results of fitting Model 3a to Hourly EP. The histogram of deviation of expected values (defined in (5.6)) is very similar to the plot of the theoretical density function. But not all the estimates are at the peak of the curves. Here, the curve of the log-likelihood function versus the correlation coefficient (ρ) is quite flat and the estimate is not at the peak.

| | Estimate | Std. | T-ratio |
|----------------|-------------|--------|---------|
| κ | 0.1140 | 0.0022 | 51.8182 |
| α | -0.5994 | 0.0737 | -8.1330 |
| ω_u | 0.6420 | 0.0237 | 27.0886 |
| γ_u | 0.3459 | 0.0092 | 37.5978 |
| ω_d | 0.4187 | 0.0338 | 12.3876 |
| γ_d | -0.3670 | 0.0151 | 24.3046 |
| ρ | -0.9987 | 0.0002 | 4993.5 |
| κ_v | 0.6915 | 0.0349 | 19.8138 |
| α_v | 0.0281 | 0.0022 | 12.7727 |
| σ_v | 0.1782 | 0.0219 | 8.1370 |
| Log-likelihood | -11343.3605 | | |

Table 5.10: Hourly EP parameters values for Model 3b

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 3b

$$\left(\begin{array}{l} d \begin{bmatrix} X_t \\ V_t \end{bmatrix} = \begin{bmatrix} \kappa(\alpha - X_t) \\ \kappa_v(\alpha_v - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{1-\rho^2}\sqrt{V_t} & \rho\sqrt{V_t} \\ 0 & \sigma_v\sqrt{V_t} \end{bmatrix} \begin{bmatrix} dW_t \\ dW_v \end{bmatrix} \\ + \begin{bmatrix} Q_t^u dP_t(\omega_u) + Q_t^d dP_t(\omega_d) \\ 0 \end{bmatrix} \end{array} \right) \text{ to Hourly EP.}$$

All the estimates are hourly based. The standard errors are still reasonably small. All the absolute values of the t-ratios are larger than 1.96, which indicates that all the parameters are significant.

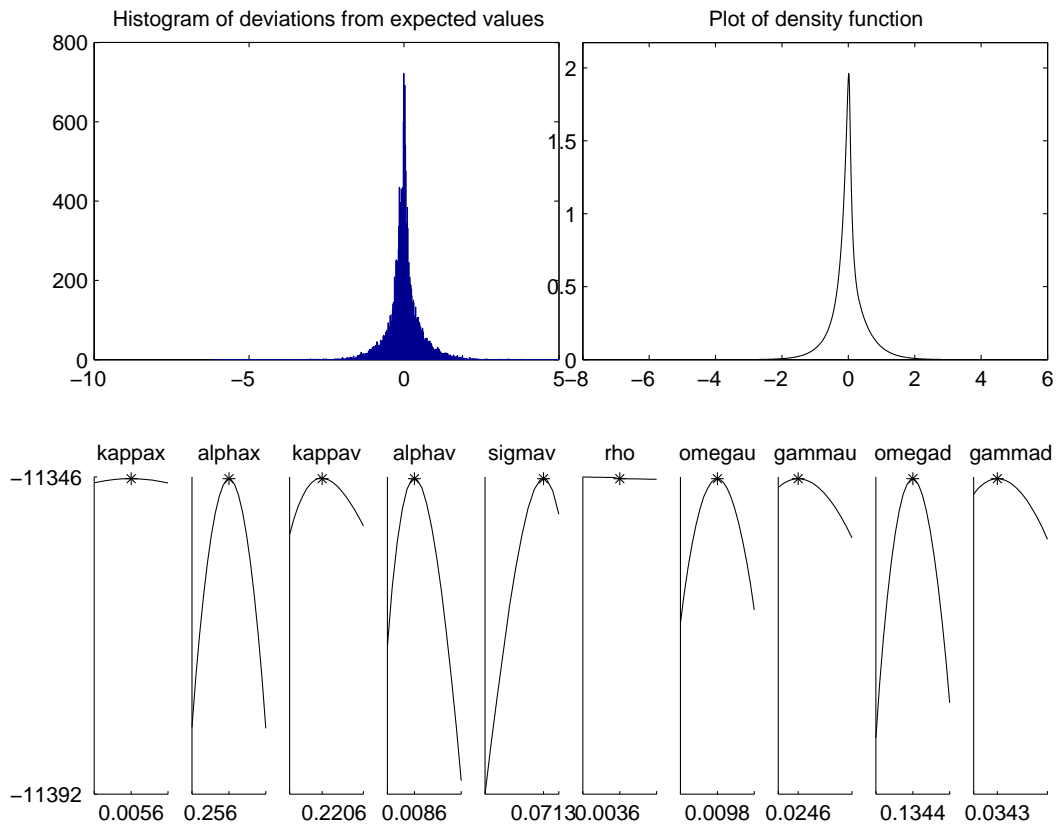


Figure 5.8: Results from optimization (Model 3b)

Note: these are the results of fitting Model 3b to Hourly EP. The histogram of deviation of expected values (defined in (5.6)) is very similar to the plot of the theoretical density function. But not all the estimates are at the peak of the curves. Similar to Model 3a, the curve of the log-likelihood function versus the correlation coefficient (ρ) is quite flat and the estimate is not at the peak.

| | True Theta | Mean | Bias | RMSE |
|------------|------------|---------|---------|--------|
| κ | 0.1148 | 0.1217 | -0.0069 | 0.0079 |
| α | -0.5169 | -0.8895 | 0.3726 | 0.9010 |
| ω | 1.0715 | 1.3234 | -0.2519 | 1.0728 |
| ψ | 0.5795 | 0.6396 | -0.0601 | 0.1529 |
| γ | 0.3532 | 0.4019 | -0.0487 | 0.1643 |
| ρ | -0.9680 | -0.5994 | -0.3686 | 0.8381 |
| κ_v | 0.6070 | 0.6088 | -0.0018 | 1.4249 |
| α_v | 0.5000 | 0.5640 | -0.0640 | 0.1487 |
| σ_v | 0.1695 | 0.0674 | 0.1021 | 0.1567 |

Table 5.11: Bias for Model 3a

5.3 Goodness of Fit

Because Model 1a is subsumed by Model 2a, likelihood ratio tests (LRT) may be applied to compare these models. Likewise, we can apply LRT to compare Model 1b and Model 2b.

LRT is a statistical test of the goodness-of-fit between two models. A relatively more complex model is compared to a simpler model to see if it fits a particular data set significantly better. If so, the additional parameters of the more complex model are often used in subsequent analyses.¹⁰ LRT is only valid if used to compare hierarchically nested models. That is, the more complex model must only differ from the simpler model by the addition of one or more parameters. Adding additional parameters will always result in a higher likelihood. However, one should also consider whether adding additional parameters results in significant improvement fitting a model to a particular data set. LRT provides one objective criterion for selecting among possible models. LRT begins with a comparison of the likelihood values of

¹⁰LRT is explained in detail by Felsenstein (1981).

| | True Theta | Mean | Bias | RMSE |
|------------|------------|---------|---------|--------|
| κ | 0.1140 | 0.1207 | -0.0067 | 0.0078 |
| α | -0.5994 | -1.4925 | 0.8931 | 0.7309 |
| ω_u | 0.6420 | 1.0963 | -0.4543 | 0.0141 |
| γ_u | 0.3459 | 0.3563 | -0.0104 | 0.0991 |
| ω_d | 0.4187 | 0.4269 | -0.0082 | 0.1891 |
| γ_d | -0.3670 | -0.4491 | -0.0821 | 0.1835 |
| ρ | -0.9987 | -0.9833 | -0.0154 | 0.1679 |
| κ_v | 0.6915 | 1.0192 | -0.3277 | 0.1272 |
| α_v | 0.5000 | 0.5404 | -0.0404 | 0.1891 |
| σ_v | 0.1782 | 0.0970 | 0.0812 | 0.1679 |

Table 5.12: Bias for Model 3b

the two models:

$$LR = 2(\ln(L_f) - \ln(L_r)), \quad (5.7)$$

where L_f and L_r are the likelihood values of the more complex model and the simpler model respectively. This LRT statistic approximately follows a chi-square distribution with degrees of freedom k , where k is equal to the number of additional parameters in the more complex model. Using this information we can then determine the critical value of the test statistic from standard statistical tables.

Table 5.15 lists the LR statistics for different pairs of models. In this table, the first column lists the pair of models under consideration. The second column labeled DF lists the additional degrees of freedom contained in the more complex model. The column labeled LR contains the actual likelihood ratio statistic. The column labeled Cutoff contains the 95% chi-squared cutoff value for a likelihood ratio statistic with degrees of freedom corresponding to the number in the DF column. The hypothesis, that the restrictions included in the simpler model are valid, is

| | Estimate | Std. | T-ratio |
|----------------|-------------|--------|---------|
| κ | 0.1135 | 0.1322 | 0.85855 |
| α | -0.5250 | 1.0000 | -0.5250 |
| ω | 1.0878 | 0.4979 | 2.1848 |
| ψ | 0.5796 | 0.4993 | 1.1608 |
| γ | 0.3508 | 0.3767 | 0.9312 |
| κ_v | 0.5884 | 0.0000 | 5884 |
| α_v | 0.0259 | 0.0374 | 0.6925 |
| σ_v | 0.1562 | 0.0119 | 13.1261 |
| Log-likelihood | -11344.1626 | | |

Table 5.13: Hourly EP parameters values for Model 3a ($\rho = -1$)

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 3a to Hourly EP with $\rho = -1$. All the estimates are hourly based. Although we do get a higher log-likelihood, some of the estimates are not significant any more (the absolute values of the t-ratio < 1.96). Moreover, the standard error of α is one, which suggests that the estimates may be not accurate.

| | Estimate | Std. | T-ratio |
|----------------|-------------|--------|---------|
| κ | 0.1135 | 0.0171 | 6.6501 |
| α | -0.6008 | 0.6502 | -0.9240 |
| ω_u | 0.6411 | 0.5239 | 1.2238 |
| γ_u | 0.3461 | 0.1690 | 2.0471 |
| ω_d | 0.4180 | 0.3974 | 1.0519 |
| γ_d | -0.3672 | 0.3612 | 1.0166 |
| κ_v | 0.6809 | 0.1277 | 5.3335 |
| α_v | 0.0281 | 0.0271 | 1.0372 |
| σ_v | 0.1769 | 0.1919 | 0.9221 |
| Log-likelihood | -11343.3294 | | |

Table 5.14: Hourly EP parameters values for Model 3b ($\rho = -1$)

Note: this table reports the estimates, the standard error of estimation and t-statistics from fitting Model 3b to Hourly EP with $\rho = -1$. All the estimates are hourly based. Although we do get a higher log-likelihood, some of the estimates are not significant any more (the absolute values of the t-ratio < 1.96).

| | DF | LR | Cutoff |
|-----------------------|----|----------|--------|
| Model 1a vs. Model 2a | 1 | 53.9024 | 3.84 |
| Model 1a vs. Model 3a | 2 | 156.5978 | 5.99 |
| Model 1b vs. Model 2b | 1 | 58.4908 | 3.84 |
| Model 1b vs. Model 3b | 2 | 140.0579 | 5.99 |

Table 5.15: Likelihood ratio statistics fitting on Hourly EP

Note: in this table, the first column lists the pair of models under consideration. The second column labeled DF lists the additional degrees of freedom contained in the more complex model. The column labeled LR contains the actual likelihood ratio statistic. The column labeled Cutoff contains the 95% chi-squared cutoff value for a likelihood ratio statistic with degrees of freedom corresponding to the number in the DF column. The hypothesis, that the restrictions included in the simpler model are valid, is rejected if the quantity in the LR column is greater than the quantity in the Cutoff column.

rejected if the quantity in the LR column is greater than the quantity in the Cutoff column. Thus this test suggests that the time-varying mean models provide a better fit than the ones with constant long-term mean and volatility. Moreover, whereas the two-jump version processes and the one-jump version processes are not nested, the log-likelihood values for the two-jump version processes are higher than their one-jump counterparts. Thus one would expect they will give a better fit.

Another related statistical measurement of fit is the Schwarz criterion (see [51]). This is defined as:

$$SC = -2 \ln L + k \ln n, \quad (5.8)$$

where L represents the likelihood value, k is the number of parameters estimated in the model, and n is the sample size. Models that yield a minimum value for the criterion are to be preferred. One should also consider whether adding additional parameters results in significant improvement fitting a model to a particular data set.

| | $\ln L$ | SC |
|----------|-------------|----------|
| Model 1a | -11424.1002 | 22907.53 |
| Model 2a | -11397.1492 | 22863.51 |
| Model 1b | -11413.3895 | 22895.99 |
| Model 2b | -11384.1441 | 22847.39 |

Table 5.16: Schwarz criteria statistics for Models

Note: in this table, the first column lists the model under consideration. The second column labeled $\ln L$ lists the log likelihood value. The column labeled SC reports the computed Schwarz criteria. Models that yield a minimum value for the criterion are to be preferred. From this table, Model 2b is the best model.

The model which has maximum likelihood value and an appropriate number of additional parameters will be chosen. Therefore, the Schwarz criterion accommodates the trade-off between a better fit and less parameters by penalising the model with additional parameters in the model. The Schwarz criterion values can be compared among various models as a basis for the selection of the model. Table 5.16 reports the log-likelihood values together with the Schwarz criteria for models. It appears that Model 2b better describes the dynamics of Hourly EP as it has the smallest Schwarz criterion.

While comparative tests like LRT and the Schwartz criterion offer an indication of the relative quality of each model, they do not yield an absolute assessment. We will do further evaluation on those models in the following sections.

5.4 Robustness Test

A simulation exercise was undertaken to determine how robust ML-CCF estimators performed. The row labeled True Theta is used to generate 300 sample paths from the

| | κ | α | σ^2 | ω | ψ | γ |
|------------|----------|----------|------------|----------|---------|----------|
| True Theta | 0.1124 | -0.2368 | 0.0144 | 1.2260 | 0.5261 | 0.3367 |
| Mean | 0.1194 | -0.2299 | 0.0160 | 1.2407 | 0.5249 | 0.3560 |
| Mean Bias | 0.0070 | 0.0069 | 0.0016 | 0.0147 | -0.0012 | 0.0193 |
| RMSE | 0.0095 | 0.0760 | 0.0029 | 0.0984 | 0.0165 | 0.0260 |

Table 5.17: Robustness test for Model 1a

| | κ | α | σ^2 | ω_u | γ_u | ω_d | γ_d |
|------------|----------|----------|------------|------------|------------|------------|------------|
| True Theta | 0.1162 | -0.1405 | 0.0131 | 0.6190 | 0.3536 | 0.6622 | -0.3059 |
| Mean | 0.1241 | -0.1439 | 0.0148 | 0.6252 | 0.3758 | 0.6664 | -0.3265 |
| Mean Bias | 0.0079 | -0.0034 | 0.0017 | 0.0062 | 0.0222 | 0.0042 | 0.0206 |
| RMSE | 0.0100 | 0.0871 | 0.0029 | 0.0587 | 0.0332 | 0.0689 | 0.0307 |

Table 5.18: Robustness test for Model 1b

discretized model beginning with the long-term mean α . Because each sample path is for an hourly time interval, we assume the generated sample paths are also for an hourly time interval. We use Milstein's scheme to simulate 3000 hourly observations for each path, and the 300 paths are generated using the antithetic variate technique. Estimation is undertaken via the ML-CCF estimation. The estimates are provided in Tables 5.17–5.20. We report the true parameter values, along with those statistical measures from the estimation. Notice that ML-CCF estimators performed quite well and our estimation procedure is very accurate since the mean bias and RMSE are small.

| | κ | α | σ^2 | ω | ψ | γ | α_1 |
|------------|----------|----------|------------|----------|---------|----------|------------|
| True Theta | 0.1114 | -0.2353 | 0.0151 | 1.2105 | 0.5216 | 0.3598 | -0.2254 |
| Mean | 0.1051 | -0.2765 | 0.0136 | 1.2188 | 0.5228 | 0.3384 | -0.2009 |
| Mean Bias | 0.0063 | 0.0412 | 0.0015 | -0.0083 | -0.0012 | 0.0214 | -0.0245 |
| RMSE | 0.0083 | 0.0917 | 0.0027 | 0.0865 | 0.0187 | 0.0280 | 0.0832 |

Table 5.19: Robustness test for Model 2a

| | κ | α | σ^2 | ω_u | γ_u | ω_d | γ_d | α_1 |
|------------|----------|----------|------------|------------|------------|------------|------------|------------|
| True Theta | 0.1084 | -0.1665 | 0.0121 | 0.6096 | 0.3569 | 0.6725 | -0.3043 | -0.0870 |
| Mean | 0.1151 | -0.1223 | 0.0135 | 0.6145 | 0.3788 | 0.6817 | -0.3194 | -0.1156 |
| Mean Bias | 0.0067 | 0.0443 | 0.0014 | 0.0048 | 0.0219 | 0.0092 | 0.0151 | -0.0287 |
| RMSE | 0.0087 | 0.1076 | 0.0027 | 0.0566 | 0.0329 | 0.0738 | 0.0270 | 0.1009 |

Table 5.20: Robustness test for Model 2b

5.5 Descriptive Statistics of Empirical vs. Calibrated Hourly Returns

Given the CCF, we can obtain the moments for any choice of jump distribution where the jump intensity or distribution does not depend on the state variables. To obtain the moments, we differentiate the CCF (ϕ) successively with respect to s and then find the value of the derivative when $s = 0$. Let U_n denote the n th moment, and ϕ_n be the n -th derivative of ϕ with respect to s , i.e., $\phi_n = \frac{\partial^n \phi}{\partial s^n}$. Then

$$U_n := \frac{1}{i^n} [\phi_n | s = 0]. \quad (5.9)$$

For Model 1a, the CCF of X_T given X_t has the closed form

$$\begin{aligned} \phi(s, \theta, X_T | X_t) &= E[\exp(isX_T) | X_t] \\ &= \exp(A(s, t, T, \theta) + B(s, t, T, \theta)X_t), \end{aligned} \quad (5.10)$$

where

$$\begin{aligned}
A(s, t, T, \theta) &= i\alpha s(1 - e^{-\kappa(T-t)}) - \frac{\sigma^2 s^2}{4\kappa}(1 - e^{-2\kappa(T-t)}) \\
&\quad + \frac{i\omega(1 - 2\psi)}{\kappa}(\arctan(\gamma s e^{-\kappa(T-t)}) - \arctan(\gamma s)) \\
&\quad + \frac{\omega}{2\kappa} \ln\left(\frac{1 + \gamma^2 s^2 e^{-2\kappa(T-t)}}{1 + \gamma^2 s^2}\right),
\end{aligned} \tag{5.11}$$

$$B(s, t, T, \theta) = i s e^{-\kappa(T-t)}.$$

Then the unconditional (stationary) characteristic function $\phi(s, \theta)$ of X_T , which is obtained by letting $T \rightarrow \infty$, is of the form

$$\phi(s, \theta) = \exp\left(i\alpha s - \frac{\sigma^2 s^2}{4\kappa} + \frac{i\omega(2\psi - 1)}{\kappa} \arctan(\gamma s) - \frac{\omega}{2\kappa} \ln(1 + \gamma^2 s^2)\right). \tag{5.12}$$

We will let X_∞ denote the random variable with this unconditional (stationary) distribution. One can obtain the n -th moment by equation (5.9). For example, to get the first moment, we differentiate $\phi(s, \theta)$ with respect to s :

$$\phi_1 := \frac{\partial \phi(s, \theta)}{\partial s} = \left(i\alpha - \frac{\sigma^2 s}{2\kappa} + \frac{i\omega(2\psi - 1)}{\kappa} \frac{\gamma}{1 + \gamma^2 s^2} - \frac{\omega}{2\kappa} \frac{2\gamma^2 s}{1 + \gamma^2 s^2}\right) \phi(s, \theta). \tag{5.13}$$

Thus the first moment U_1 is given by

$$U_1 := E[X_\infty] = \frac{1}{i}[\phi_1|s=0] = \alpha + \frac{\omega(2\psi - 1)}{\kappa} \gamma. \tag{5.14}$$

Similarly, we can obtain other moments:

$$\begin{aligned}
U_2 &:= E[X_\infty^2] = \frac{\sigma^2 + 2\omega\gamma^2}{2\kappa} + U_1^2, \\
U_3 &:= E[X_\infty^3] = \frac{2\omega(2\psi - 1)\gamma^3}{\kappa} + 3U_1U_2 - 2U_1^3, \\
U_4 &:= E[X_\infty^4] = \frac{6\omega\gamma^4}{\kappa} + 3U_2^2 - 12U_2U_1^2 + 4U_3 + 6U_1^4.
\end{aligned} \tag{5.15}$$

Therefore, we have the unconditional variance, skewness and kurtosis:

$$\begin{aligned}
\text{Variance} &:= E[(X_\infty - U_1)^2] = \frac{\sigma^2 + 2\omega\gamma^2}{2\kappa}, \\
\text{Skewness} &:= E\left[\frac{(X_\infty - U_1)^3}{(U_2 - U_1^2)^{\frac{3}{2}}}\right] = \frac{4\sqrt{2\kappa}\omega\gamma^3(2\psi - 1)}{(\sigma^2 + 2\omega\gamma^2)^{\frac{3}{2}}}, \\
\text{Kurtosis} &:= E\left[\frac{(X_\infty - U_1)^4}{(U_2 - U_1^2)^2}\right] = \frac{24\kappa(\omega\gamma^4)}{(\sigma^2 + 2\omega\gamma^2)^2} + 3.
\end{aligned} \tag{5.16}$$

We can also obtain the conditional variance, skewness and kurtosis for the process in the same fashion:

$$\begin{aligned}
\text{Variance} &:= E[(X_{t+1} - U_1)^2 | X_t] = \left(\frac{\sigma^2 + 2\omega\gamma^2}{2\kappa}\right)(1 - e^{-2\kappa}), \\
\text{Skewness} &:= E\left[\frac{(X_{t+1} - U_1)^3}{(U_2 - U_1^2)^{\frac{3}{2}}} | X_t\right] = \frac{4\omega\sqrt{2\kappa}(2\psi - 1)\gamma^3}{((\sigma^2 + 2\omega\gamma^2)(1 - e^{-2\kappa}))^{\frac{3}{2}}}(1 - e^{-3\kappa}), \\
\text{Kurtosis} &:= E\left[\frac{(X_{t+1} - U_1)^4}{(U_2 - U_1^2)^2} | X_t\right] = \frac{24\kappa\omega\gamma^4(1 + e^{-2\kappa})}{(\sigma^2 + 2\omega\gamma^2)^2(1 - e^{-2\kappa})} + 3.
\end{aligned} \tag{5.17}$$

Notice that since higher moments (variance, skewness and kurtosis) do not include the information of the long-term mean, Model 2a has the same formulae for higher moments as Model 1a.

As a first check we compute the moments of the unconditional distribution and conditional distribution of the logarithm of electricity prices, using the estimated parameters for the Model 1a and Model 2a in Table 5.5 and Table 5.6 respectively. Empirical and theoretical results are listed in Tables 5.21 and 5.22. According to Das [52], empirical approximations of the unconditional moments can be obtained by computing higher moments of the empirical data (Hourly Peak), and the empirical conditional moments are approximated by computing higher moments of the changes of the empirical data (Hourly Peak).

For Model 1b, we can obtain the unconditional characteristic function in the same

| | Empirical results | | Theoretical results | |
|----------|-------------------|-------------|---------------------|-------------|
| | unconditional | conditional | unconditional | conditional |
| Std.Dev | 0.8359 | 0.5529 | 1.1404 | 0.5117 |
| Skewness | 0.1089 | -0.0133 | 0.0293 | 0.0929 |
| Kurtosis | 4.1445 | 8.4214 | 3.4972 | 7.4424 |

Table 5.21: Empirical results vs. theoretical results for Model 1a

| | Empirical results | | Theoretical results | |
|----------|-------------------|-------------|---------------------|-------------|
| | unconditional | conditional | unconditional | conditional |
| Std.Dev | 0.8359 | 0.5529 | 1.1801 | 0.5138 |
| Skewness | 0.1089 | -0.0133 | 0.0249 | 0.0817 |
| Kurtosis | 4.1445 | 8.4214 | 3.4704 | 7.4926 |

Table 5.22: Empirical results vs. theoretical results for Model 2a

fashion,

$$\phi(s, \theta) = \exp\left(i\alpha s - \frac{\sigma^2 s^2}{4\kappa} - \frac{\omega_u}{\kappa} \ln(1 - is\gamma_u) - \frac{\omega_d}{\kappa} \ln(1 - is\gamma_d)\right). \quad (5.18)$$

Thus we have the unconditional variance, skewness and kurtosis:

$$\begin{aligned} \text{Variance} &:= E[(X_\infty - U_1)^2] = \frac{\sigma^2 + 2\omega_u\gamma_u^2 + 2\omega_d\gamma_d^2}{2\kappa}, \\ \text{Skewness} &:= E\left[\frac{(X_\infty - U_1)^3}{(U_2 - U_1^2)^{\frac{3}{2}}}\right] = \frac{4\sqrt{2\kappa}(\omega_u\gamma_u^3 + \omega_d\gamma_d^3)}{(\sigma^2 + 2\omega_u\gamma_u^2 + 2\omega_d\gamma_d^2)^{\frac{3}{2}}}, \\ \text{Kurtosis} &:= E\left[\frac{(X_\infty - U_1)^4}{(U_2 - U_1^2)^2}\right] = \frac{24\kappa(\omega_u\gamma_u^4 + \omega_d\gamma_d^4)}{(\sigma^2 + 2\omega_u\gamma_u^2 + 2\omega_d\gamma_d^2)^2} + 3, \end{aligned} \quad (5.19)$$

and the conditional variance, skewness and kurtosis:

$$\begin{aligned} \text{Variance} &:= E[(X_{t+1} - U_1)^2 | X_t] = \left(\frac{\sigma^2 + 2\omega_u\gamma_u^2 + 2\omega_d\gamma_d^2}{2\kappa}\right)(1 - e^{-2\kappa}), \\ \text{Skewness} &:= E\left[\frac{(X_{t+1} - U_1)^3}{(U_2 - U_1^2)^{\frac{3}{2}}} | X_t\right] = \frac{4\sqrt{2\kappa}(\omega_u\gamma_u^3 + \omega_d\gamma_d^3)}{((\sigma^2 + 2\omega_u\gamma_u^2 + 2\omega_d\gamma_d^2)(1 - e^{-2\kappa}))^{\frac{3}{2}}}(1 - e^{-3\kappa}), \\ \text{Kurtosis} &:= E\left[\frac{(X_{t+1} - U_1)^4}{(U_2 - U_1^2)^2} | X_t\right] = \frac{24\kappa(\omega_u\gamma_u^4 + \omega_d\gamma_d^4)(1 + e^{-2\kappa})}{((\sigma^2 + 2\omega_u\gamma_u^2 + 2\omega_d\gamma_d^2)(1 - e^{-2\kappa}))^2} + 3. \end{aligned} \quad (5.20)$$

| | Empirical results | | Theoretical results | |
|----------|-------------------|-------------|---------------------|-------------|
| | unconditional | conditional | unconditional | conditional |
| Std.Dev | 0.8359 | 0.5529 | 1.1206 | 0.5523 |
| Skewness | 0.1089 | -0.0133 | 0.1029 | 0.3207 |
| Kurtosis | 4.1445 | 8.4214 | 3.5068 | 7.3809 |

Table 5.23: Empirical results vs. theoretical results for Model 1b

| | Empirical results | | Theoretical results | |
|----------|-------------------|-------------|---------------------|-------------|
| | unconditional | conditional | unconditional | conditional |
| Std.Dev | 0.8359 | 0.5529 | 1.1604 | 0.5544 |
| Skewness | 0.1089 | -0.0133 | 0.1035 | 0.3338 |
| Kurtosis | 4.1445 | 8.4214 | 3.4779 | 7.4261 |

Table 5.24: Empirical results vs. theoretical results for Model 2b

Empirical results and theoretical results are listed in Table 5.23. Similarly, Model 2b has the same formulae for higher moments. We report empirical results and theoretical results for this model in Table 5.24. Notice that theoretical moments of the two-jump version models match empirical results better than the one-jump version models. But we did not get any useful information about the comparison between Model 1b and Model 2b through computing empirical moments and theoretical moments because the results are quite close to each other.

5.6 Quantile and Quantile plot

In this section, we will apply the quantile-quantile (QQ) plot to compare Model 1b and Model 2b. A QQ plot is a graphical technique for determining if two data

sets come from the same distribution. A QQ plot is a plot of the *quantiles*¹¹ of the first data set against the quantiles of the second data set superimposed by a 45-degree reference line. If the two sets come from a common distribution, the points should fall approximately along this reference line. The greater the departure from this reference line, the greater the evidence for the conclusion that the two data sets come from populations with different distributions. We can use a QQ plot to examine whether the one-jump version models or the two-jump version models match the prices dynamics better.

Let us consider

$$Y_t := X_t - X_{t-1} + \kappa X_{t-1} \Delta t \quad (5.21)$$

with step size Δt where X_t is the log price at time t , and $\Delta t = 1$ for the hourly data series. For one-jump version models, one may generate random paths from the following model (Equation (5.22)) with the estimates in Table 5.5.

$$Y_t = \kappa \alpha \Delta t + \sigma dW_t + Q dP(\omega), \quad (5.22)$$

with the long-term mean α different for Model 1a and Model 2a.¹² Likewise, for two-jump version models, one may generate random paths from the following model (5.23) with the estimates in Table 5.7:

$$Y_t = \kappa \alpha \Delta t + \sigma dW_t + Q_u dP_u(\omega_u) + Q_d dP_d(\omega_d), \quad (5.23)$$

with the long-term mean α different for Model 1b and Model 2b.¹³ Furthermore, we can obtain a sample Y from the historical data.

¹¹By a quantile, we mean the fraction (or percent) of points below a given value. That is, the 0.3 (or 30%) quantile is the point at which 30% of the data falls below and 70% falls above that value.

¹²Refer back to Equation (3.6) and Equation (3.20). By doing so, we obtain the stationary process (5.22).

¹³Refer back to (3.13) and (3.29). By doing so, we obtain the stationary process (5.23).

Figure 5.9 illustrates QQ plots for the different jump version data versus the adjusted historical data. The horizontal axis shows the quantiles computed from the adjusted historical data by (5.21) while the vertical axis shows the quantiles for one typical simulated path. The points in each QQ plot are close to the reference line, which indicates that the performance of all the models are satisfactory. But it is hard to tell which model is better from the QQ plots since they are quite similar.

5.7 Simulation Study

Using the parameters we obtained in Tables 5.5–5.8, we can simulate the hourly price series to compare with Hourly EP. Here we display the sample plot of one typical simulated path (dashed line), and the sample plot of Hourly EP (solid line) in Figures 5.10–5.13. The 95% quantile plot and 5% quantile plot are also shown in the same figure. Notice that most of the empirical data except for those extreme “spikes” are inside the simulation range. In this case, it is hard to tell the performance of which model is the best. Furthermore, the histogram of the changes of Hourly EP, and the corresponding distribution of the log returns of the simulated data (an overlaid black line) are shown in Figures 5.14–5.17. All the models underestimate the number of small changes.

If we separate the on-peak/off-peak data, then one only need to fit Model 1a and Model 1b to the data and compare these two models. Using the resulting estimates from fitting Peak EP and Offpeak EP, we can simulate price series to compare with the empirical data. We display plots of one typical simulated path (dashed line), and the plot of the actual data (solid line) in Figures 5.18–5.21. The 95% quantile plot

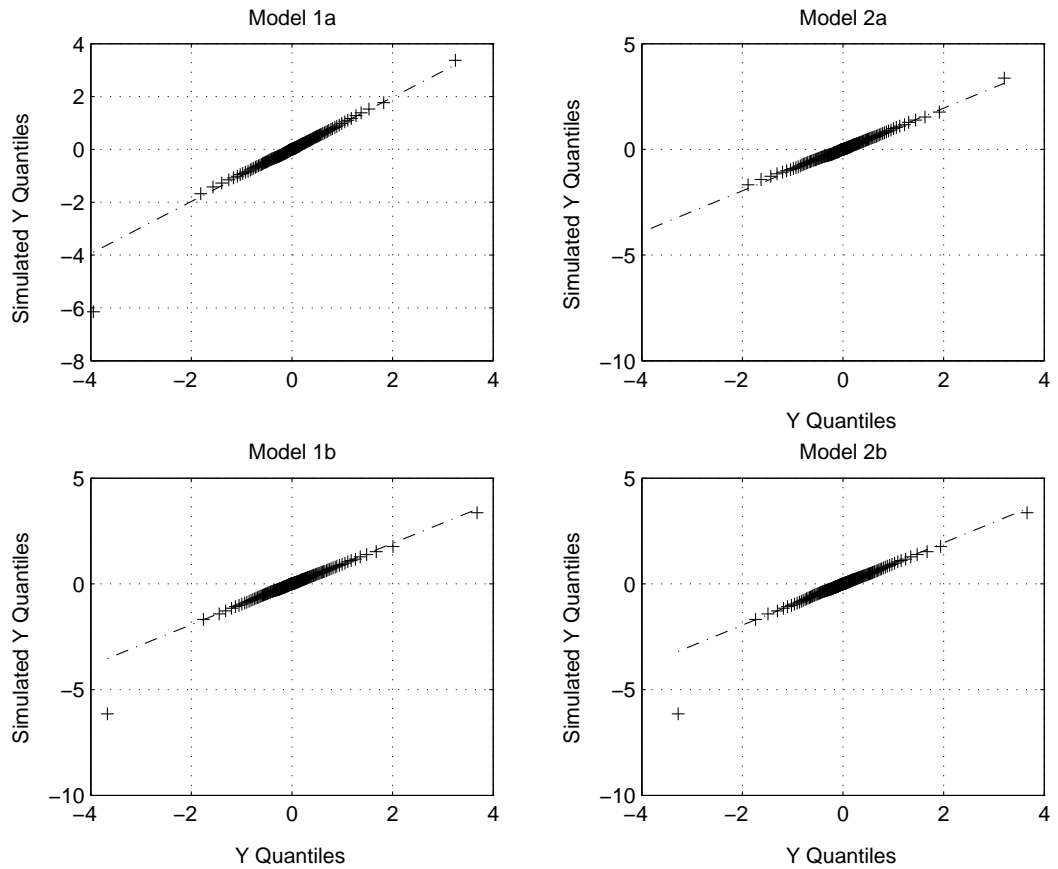


Figure 5.9: QQ Plots

Note: in each QQ plot, the horizontal axis shows the quantiles for the adjusted historical data. The vertical axis shows the quantiles for one typical simulated path for the model whose name is listed in the subtitle.

| | onpeak | Model 1a | Model 1b |
|----------|---------|----------|----------|
| Mean | 4.07631 | 4.04856 | 4.08371 |
| Std.Dev | 0.53402 | 0.67293 | 0.63997 |
| Skewness | 0.13465 | 0.31442 | 0.48105 |
| Kurtosis | 3.65594 | 4.06480 | 3.89188 |

Table 5.25: Descriptive statistics of Peak EP and simulated paths

and 5% quantile plot are also plotted in the same figure for comparison. Still most of the empirical data except for those extreme “spikes” are inside the simulation range. The histogram of the actual data and the corresponding distribution of the log returns of the simulated data (an overlaid black line) are shown in Figures 5.22–5.25. Moreover, Table 5.25 and Table 5.26 list the comparison of the moments of the actual data and the simulated data from Model 1a and Model 1b. It seems that Model 1b does a better job since the moments are closer to the empirical data.

5.8 Conclusion

This chapter report comparisons among all the models via different means. A simulation study shows that ML-CCF estimators are very robust and the estimation procedures are quite accurate. But we failed to apply ML-MCCF estimators in stochastic volatility models and we did not get accurate results from SGMM estimation in the time constraint. Therefore, a comparison between stochastic volatility models and the other four models (deterministic volatility) are not supplied. The deterministic volatility models all performed quite well in various comparisons. There is only a little bit improvement of the two-jump version models over the one-jump

| | offpeak | Model 1a | Model 1b |
|----------|---------|----------|----------|
| Mean | 3.46474 | 3.42967 | 3.47128 |
| Std.Dev | 0.64445 | 0.75322 | 0.69549 |
| Skewness | 0.21882 | 0.31544 | 0.20126 |
| Kurtosis | 2.76840 | 3.13474 | 3.16340 |

Table 5.26: Descriptive statistics of Peak EP and simulated paths

version models. As the only extra term in Model 2b is the onpeak/offpeak dummy variables, we feel that, with proper data preparation (such as deseasonalization, and separation of the on-peak data and the off-peak data), Model 1b will suffice in mimicking the dynamics in electricity price processes of Alberta.

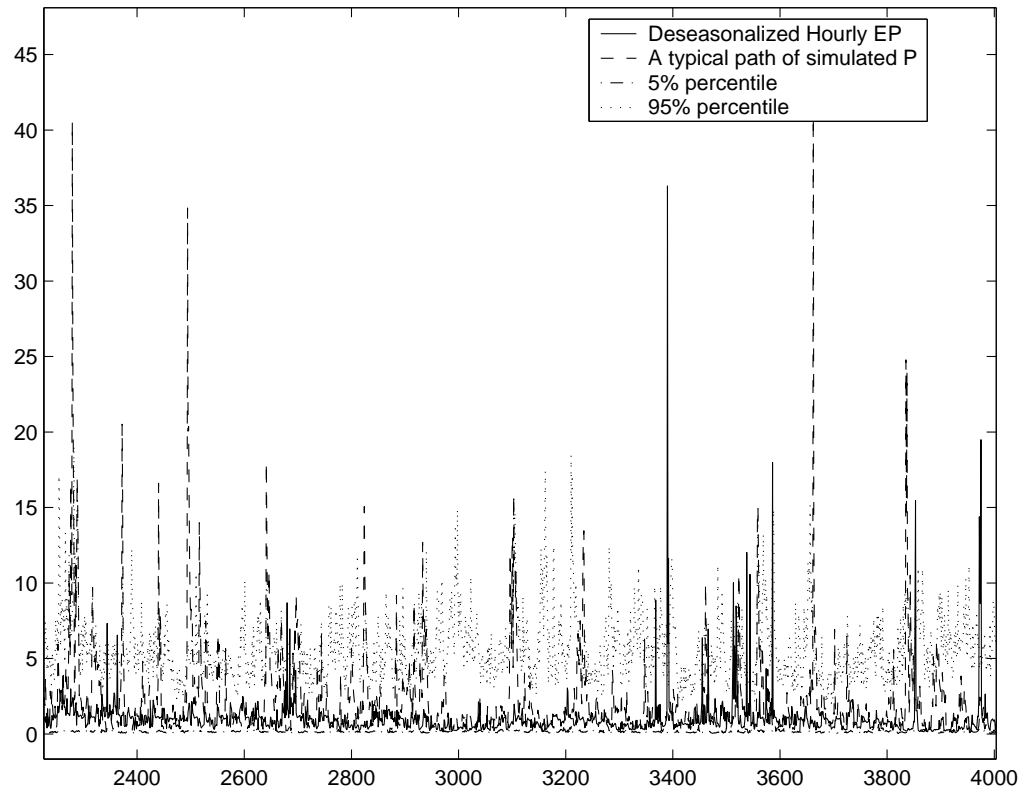


Figure 5.10: Hourly EP superimposed by simulated paths (Model 1a)
Note: this figure includes the sample plot of the deseasonalised Hourly EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

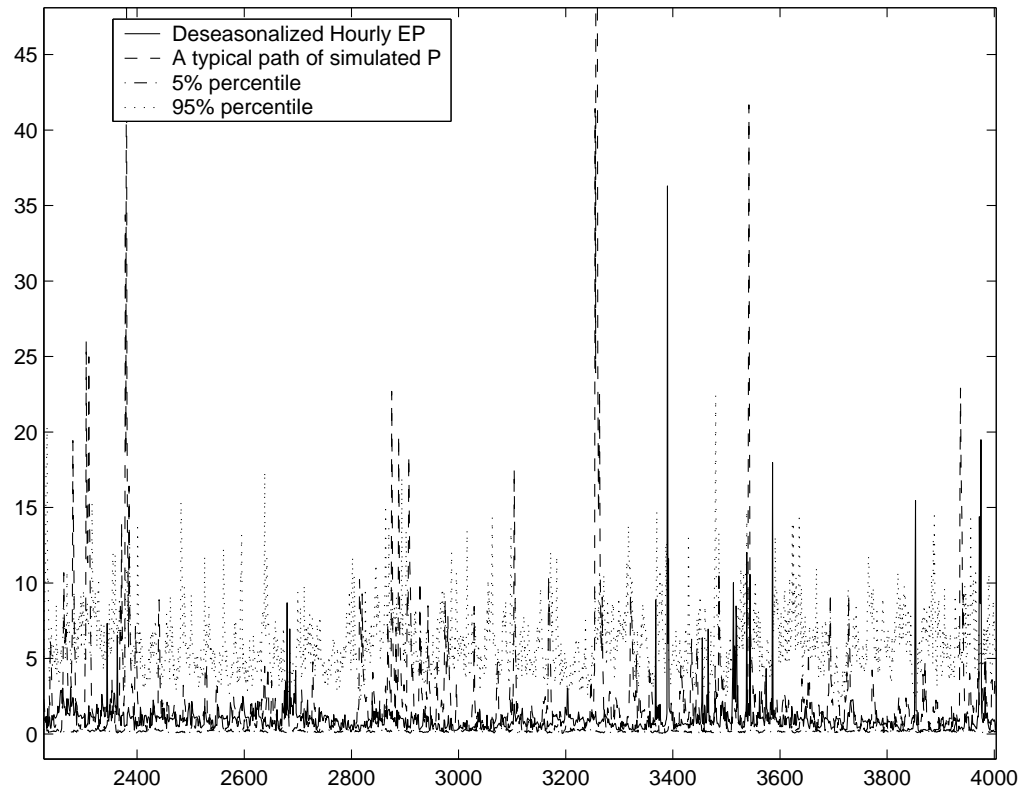


Figure 5.11: Hourly EP superimposed by simulated paths (Model 1b)
Note: this figure includes the sample plot of the deseasonalied Hourly EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

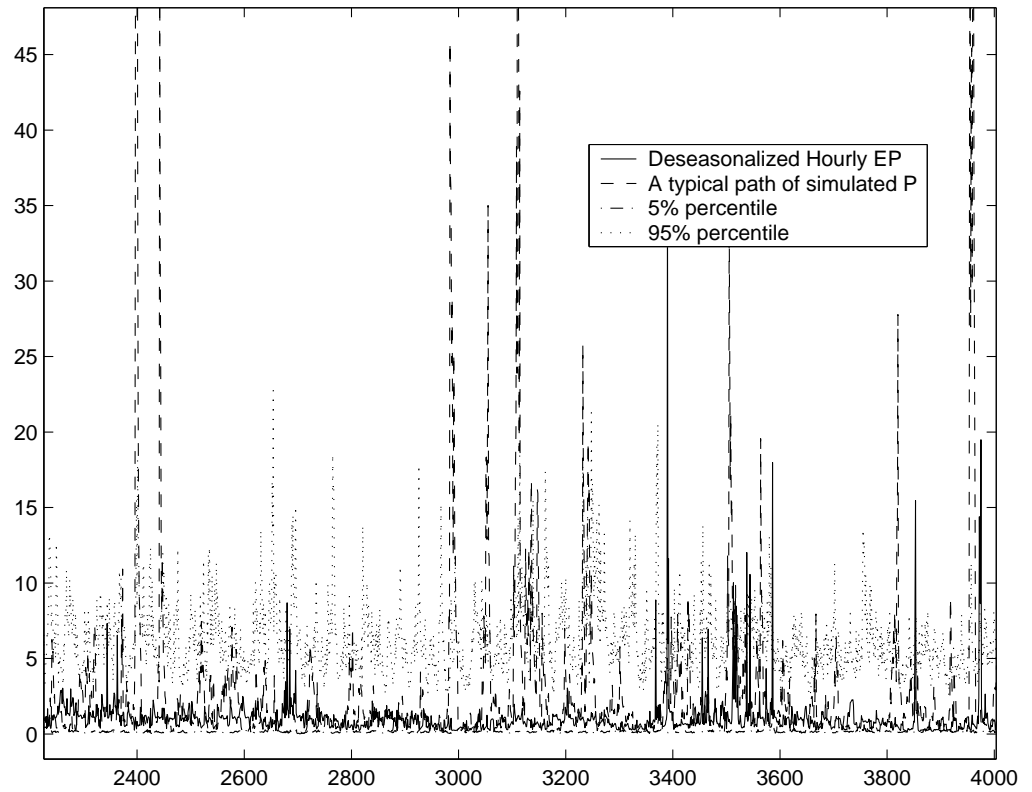


Figure 5.12: Hourly EP superimposed by simulated paths (Model 2a)
Note: this figure includes the sample plot of the deseasonalised Hourly EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

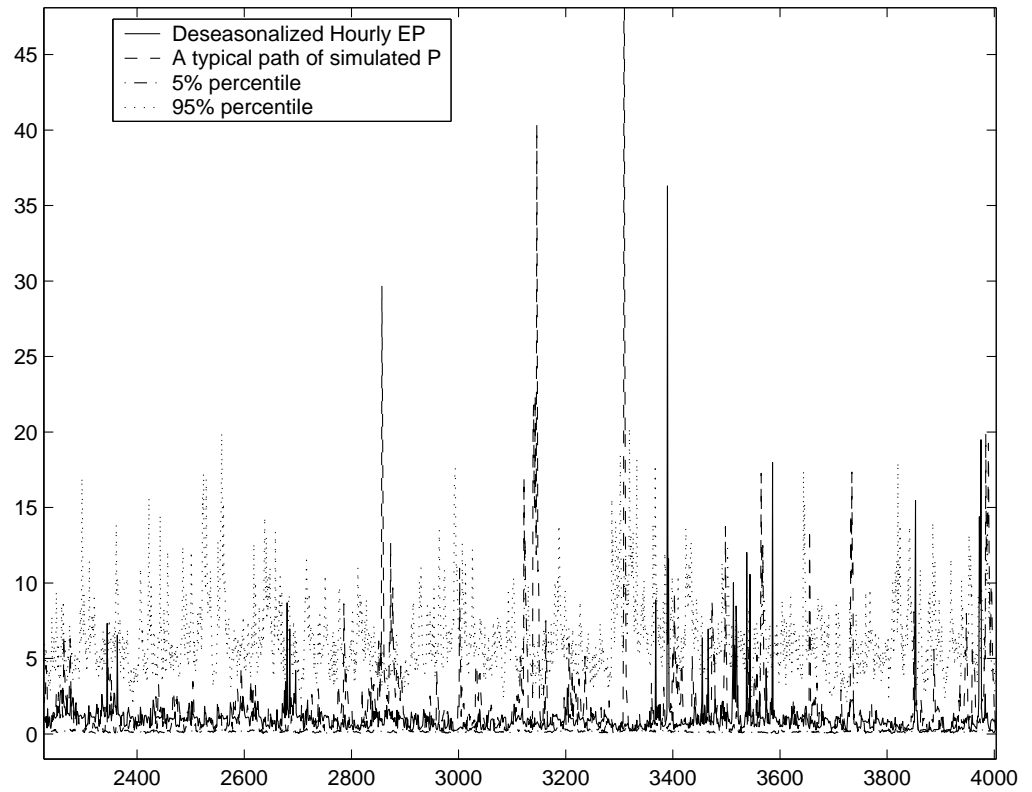


Figure 5.13: Hourly EP superimposed by simulated paths (Model 2b)
Note: this figure includes the sample plot of the deseasonalied Hourly EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

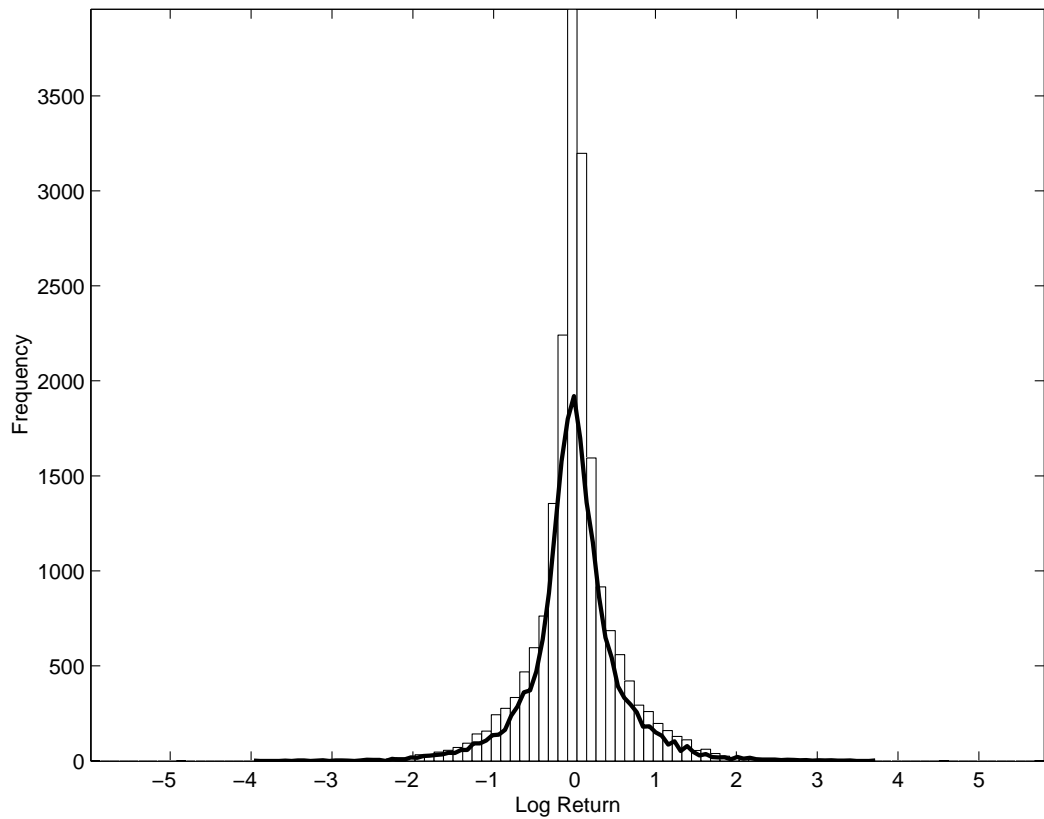


Figure 5.14: Comparison of simulated price processes with Hourly EP (Model 1a)
Note: the histogram is of the change of the deseasonalized Hourly EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes.

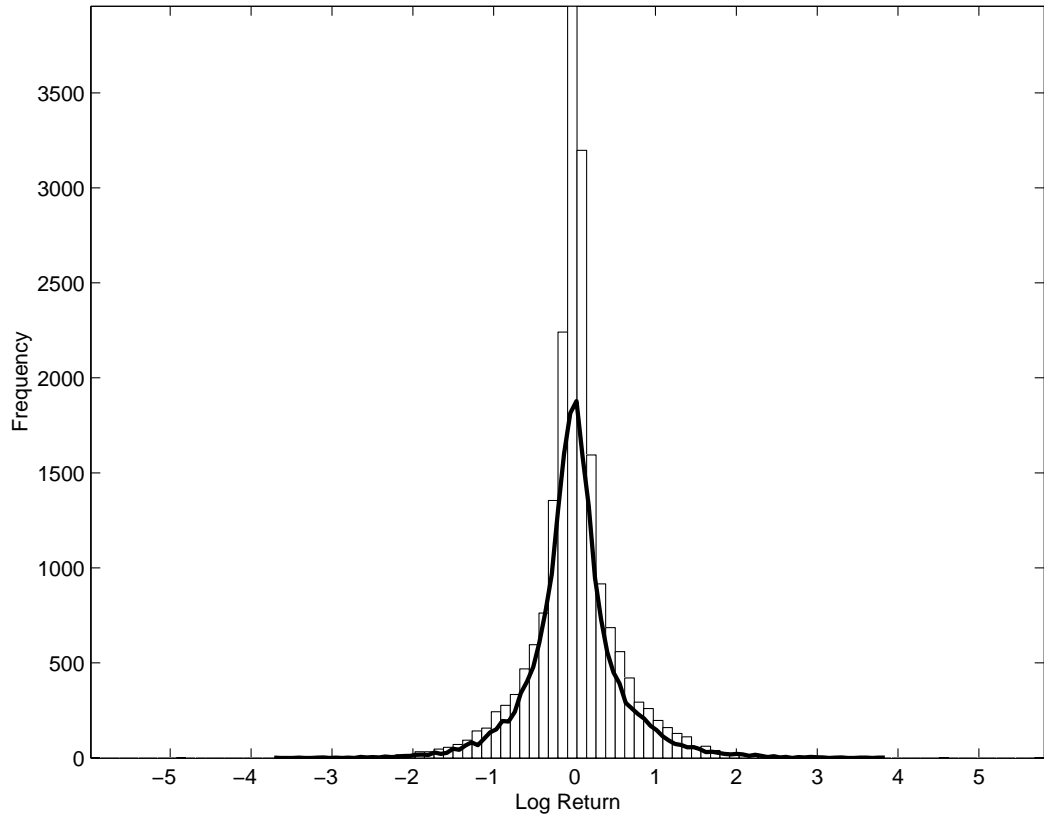


Figure 5.15: Comparison of simulated price processes with Hourly EP (Model 1b)
Note: the histogram is of the change of the deseasonalized Hourly EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes.

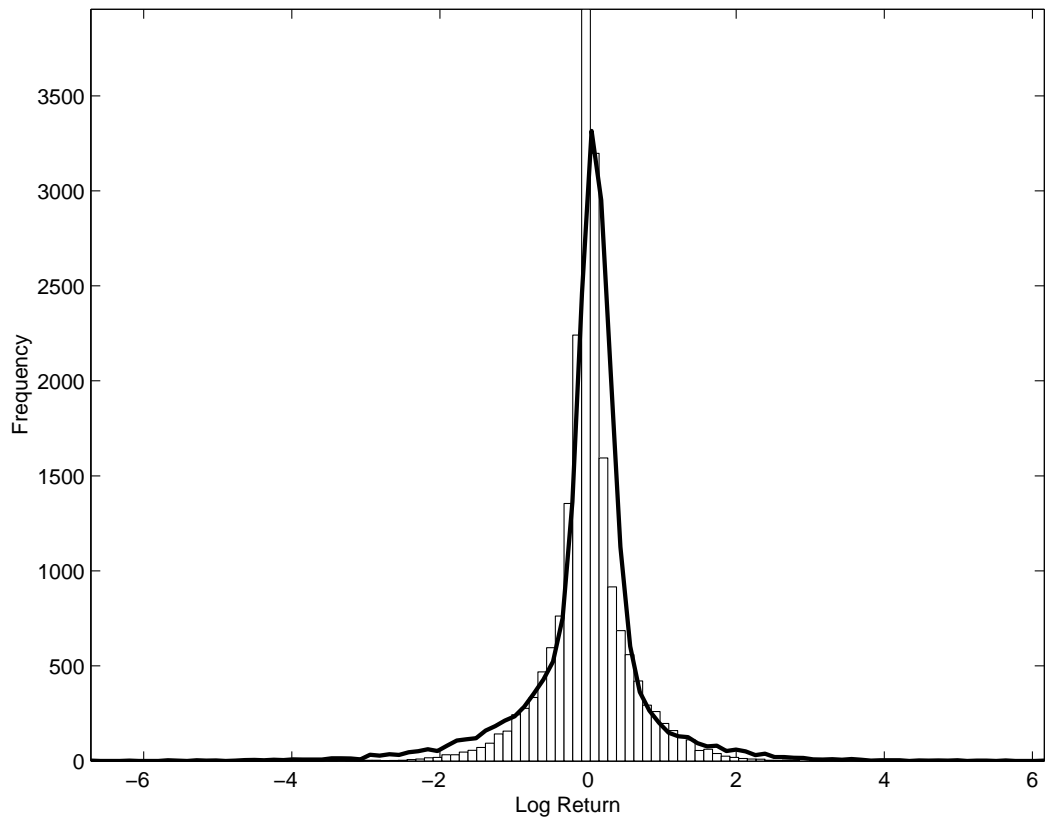


Figure 5.16: Comparison of simulated price processes with Hourly EP (Model 2a)
Note: the histogram is of the change of the deseasonalized Hourly EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes.

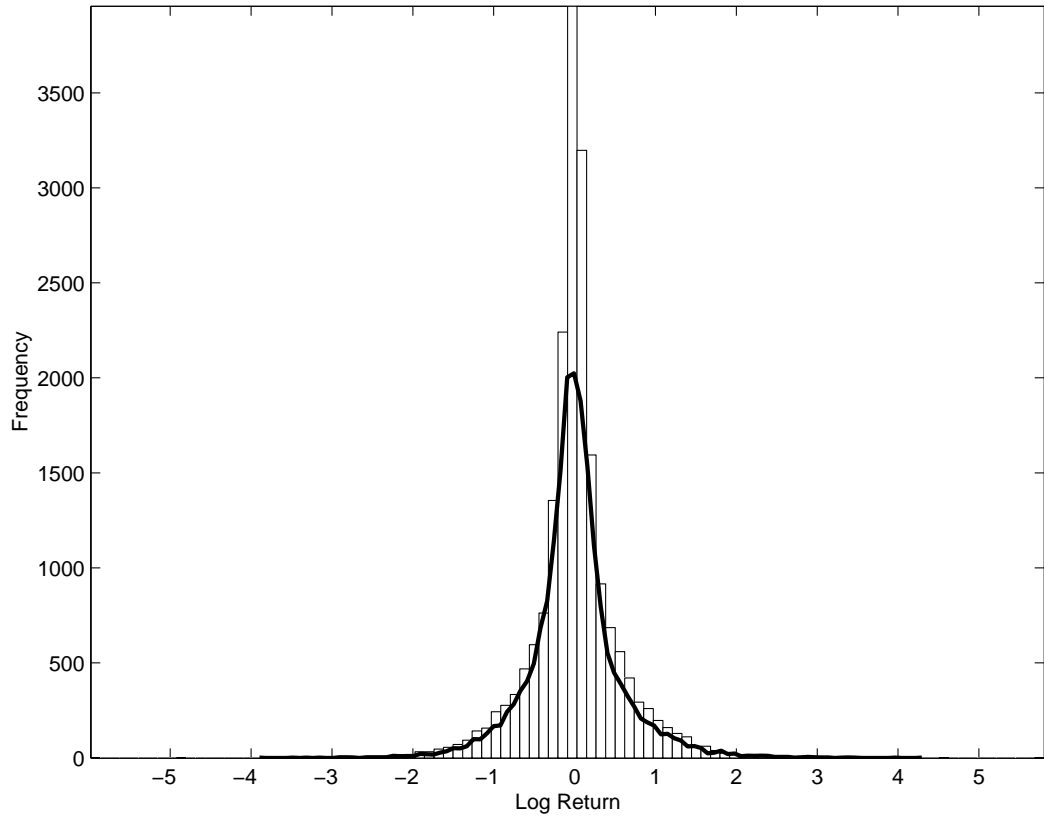


Figure 5.17: Comparison of simulated price processes with Hourly EP (Model 2b)
Note: the histogram is of the change of the deseasonalized Hourly EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes.

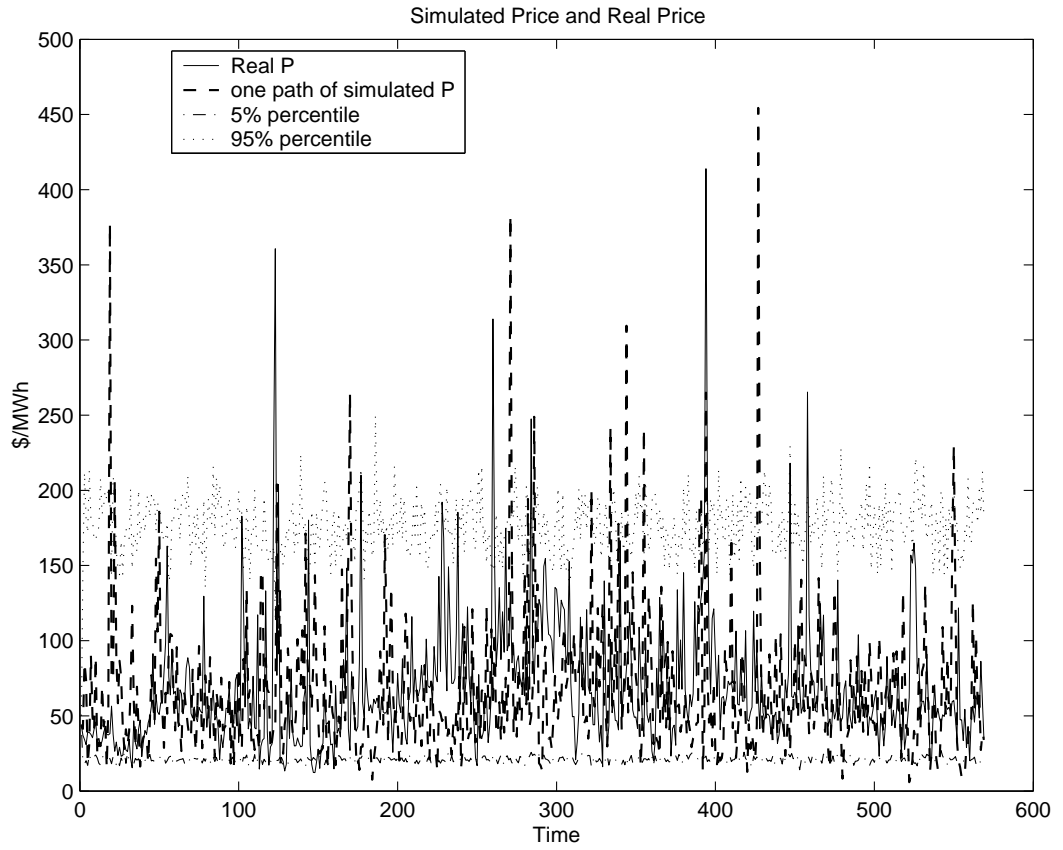


Figure 5.18: Peak EP superimposed by simulated paths (Model 1a)

Note: this figure includes the sample plot of Peak EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

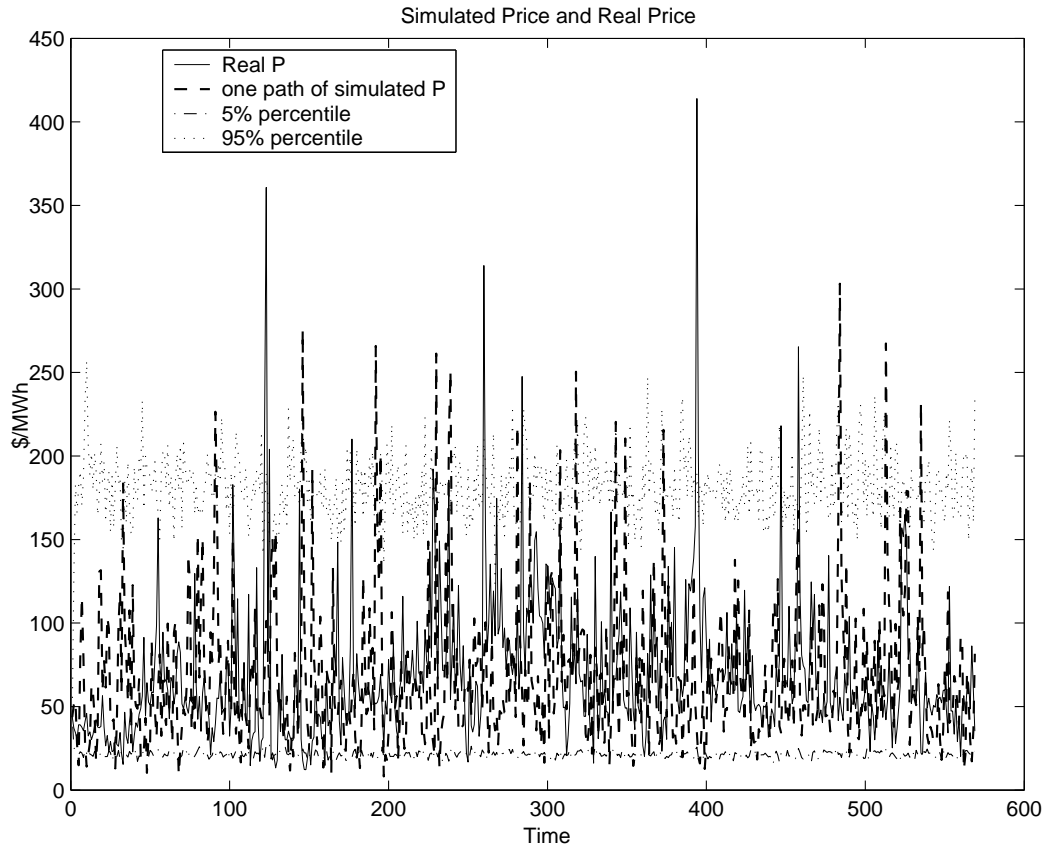


Figure 5.19: Peak EP superimposed by simulation paths (Model 1b)

Note: this figure includes the sample plot of Peak EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

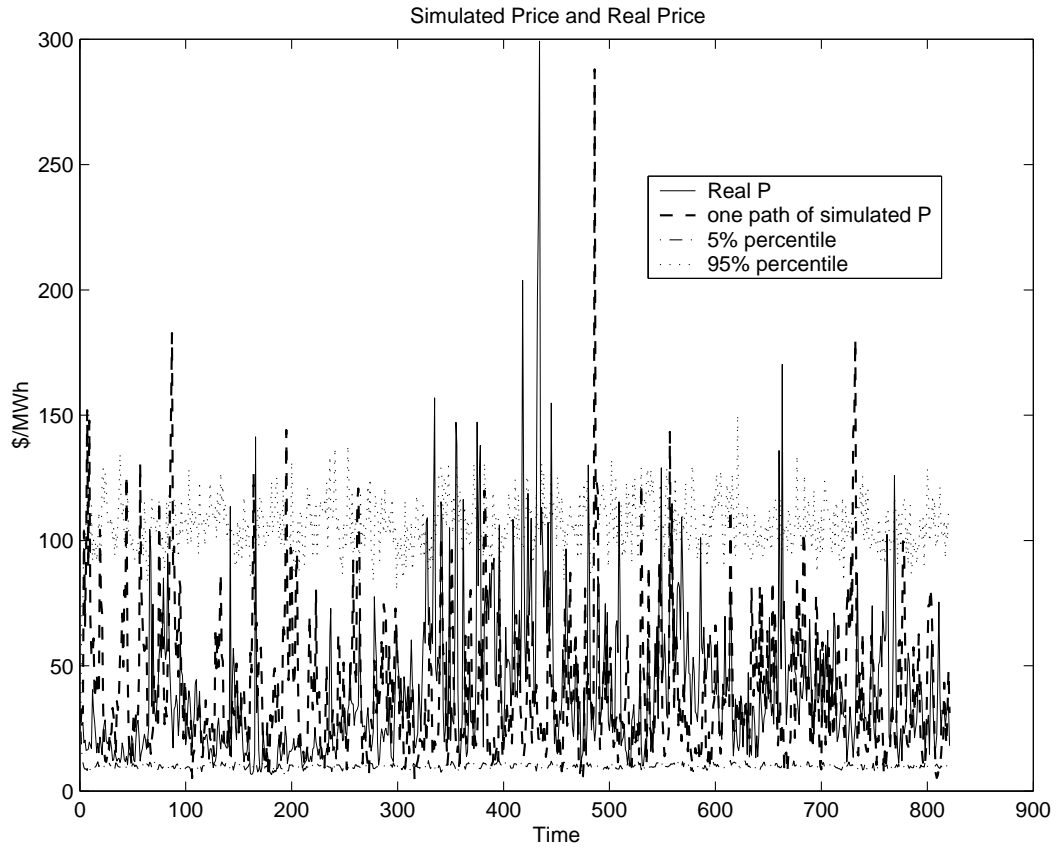


Figure 5.20: Offpeak EP superimposed by simulation paths (Model 1a)

Note: this figure includes the sample plot of Offpeak EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

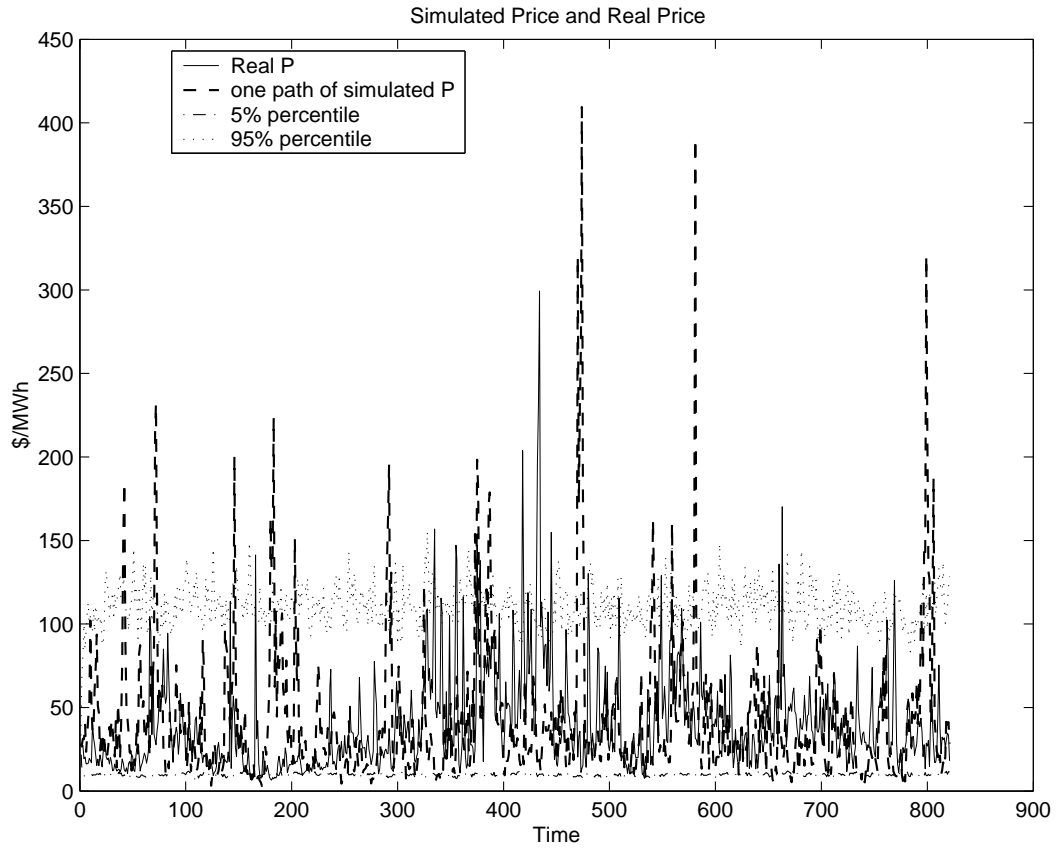


Figure 5.21: Offpeak EP superimposed by simulation paths (Model 1b)

Note: this figure includes the sample plot of Offpeak EP, one typical simulated path, the 95% quantile, and 5% quantile of the simulation.

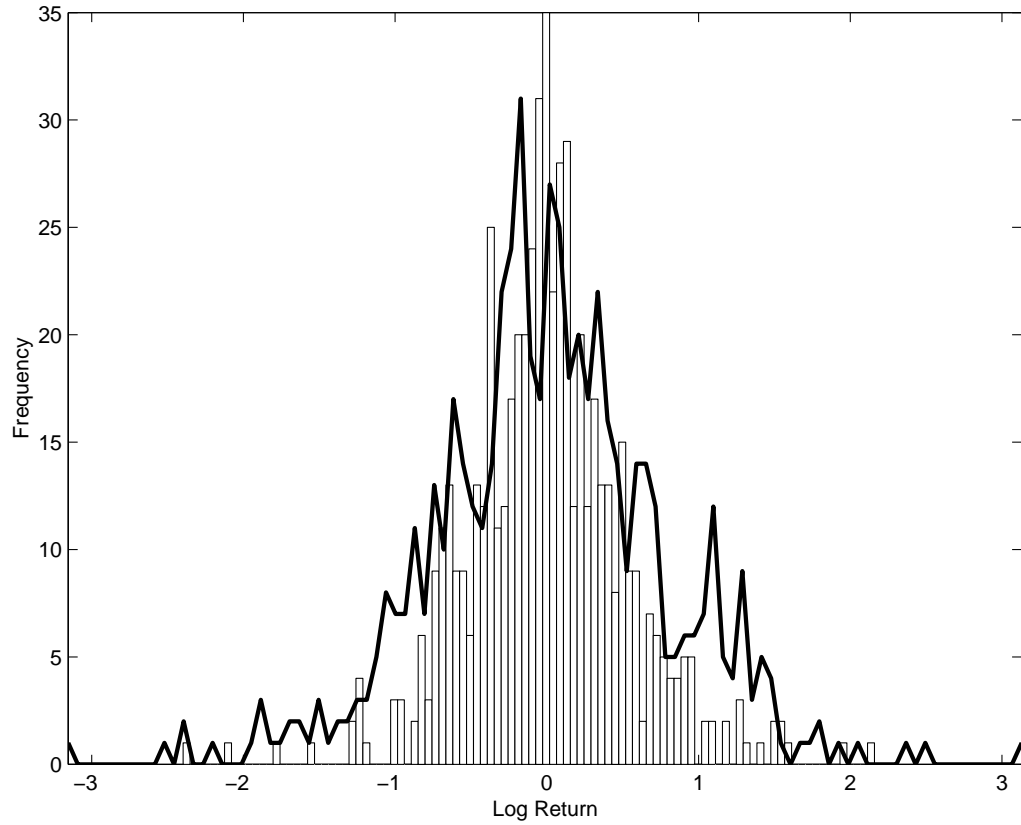


Figure 5.22: Comparison of simulated price processes with Peak EP (Model 1a)
 Note: the histogram is of the change of the deseasonalized Peak EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model overestimates the number of medium-sized changes.

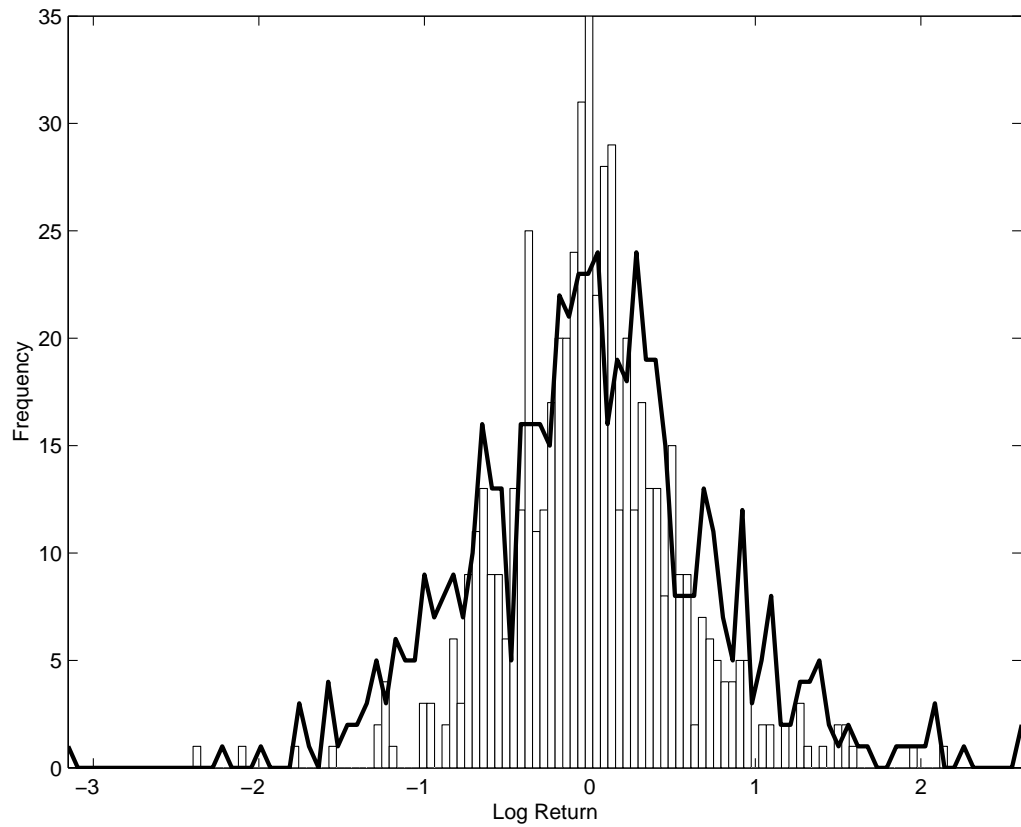


Figure 5.23: Comparison of simulated price processes with Peak EP (Model 1b)
 Note: the histogram is of the change of the deseasonalized Peak EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes and overestimates the medium-sized changes.

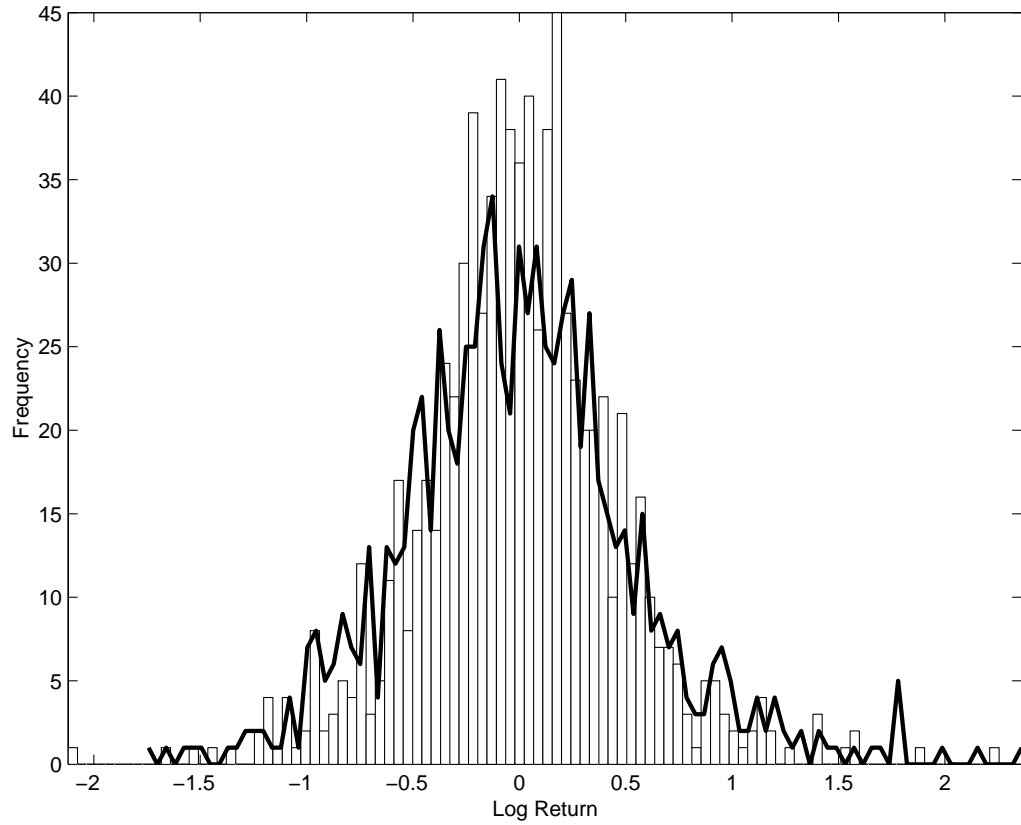


Figure 5.24: Comparison of simulated price processes with Offpeak EP (Model 1a)
 Note: the histogram is of the change of the deseasonalized Offpeak EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes.

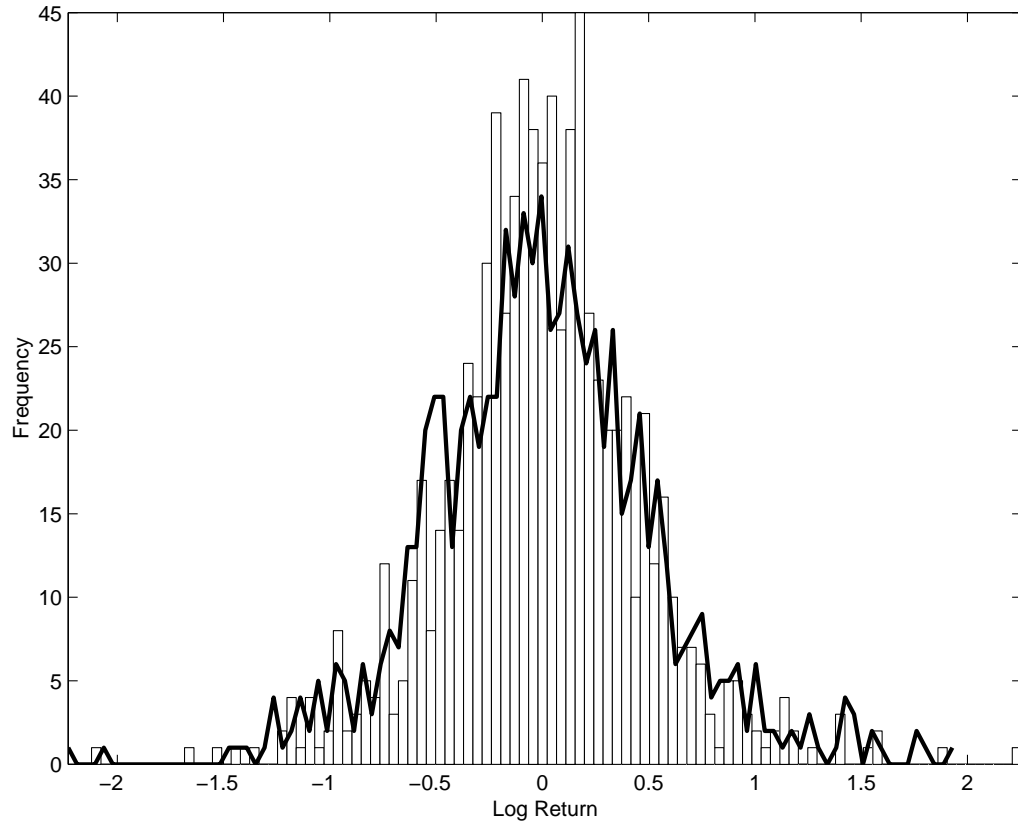


Figure 5.25: Comparison of simulated price processes with Offpeak EP (Model 1b)
Note: the histogram is of the change of the deseasonalized Offpeak EP and the overlaid black line is the corresponding distribution of the log returns of one typical simulated path. Note that this model underestimates the number of small changes.

Chapter 6

Conclusion

Energy commodity markets are growing rapidly as the deregulation and restructuring of electricity furnish industries spreads in North America and around the world. There is a heightened awareness of the need to understand the dynamics of electricity markets for trading electricity, risk management and project pricing.

Due to the difficulties in storing electricity, the interplay between supply and demand produces unique electricity price dynamics in each different electricity market. Thus electricity poses the biggest challenge for researchers and practitioners to model its price behaviour among all the energy commodities. Furthermore, because of the relative newness of deregulation, there have been few empirical studies focusing entirely on electricity prices.

In this thesis, we have addressed issues of modeling electricity spot prices in Alberta. We examined a broad class of stochastic models which can be used to model the behaviour of electricity pricing including mean-reversion, time-varying mean, stochastic volatility, as well as multiple jumps. We demonstrated how to get the CCF by means of the transform analysis. We also introduced methods showing how to fit those models via ML-CCF and SGMM estimation to the Alberta hourly electricity prices. Through extensive empirical comparisons among the models, we showed that the two-jump version of mean-reverting jump-diffusion (Model 1b) is generally superior to other models after the proper deseasonalization and separation of the on-peak and off-peak data.

In future research, a number of important issues ought to be considered. Affine processes can be applied to multiple types of process such as those with stochastic volatility, and jumps, and more general structures such as multiple latent variables and time-varying jump component, without sacrificing option pricing tractability. Although this thesis concentrated on calibration, they have many additional advantages that could be further investigated. In future work, we should be able to compute the prices of various electricity derivatives (options) under the assumed underlying affine jump-diffusion price processes by exploiting the transform analysis when applicable.

Stochastic volatility models should be further considered from a purely statistical perspective. They may perform better if we fully exploit the information of the joint CCF and use other estimation methodologies such as the Bayesian Monte Carlo Markov Chain, efficient method of moments, generalized method of moments, simulated method of moments, Kalman filtering methods, and others.

However, generating useful forecasts from jump-diffusion models is difficult. Averaging processes could cause the loss of much important information. As pointed out by Knittel and Roberts (see [26]), “One must simulate a forecasted path because of the models dependency on random jumps. Of course, there are a continuum of possible paths that may be simulated by, intuitively, flipping a λ -coin each time period, and drawing from the appropriate conditional distribution. Combining many simulated paths averages out the excess variation induced from the jumps, and leaves a very smooth forecast representing a number falling somewhere between the means of the two conditional distributions that make up the mixture.”

Furthermore, all the models we consider cannot capture unexpected events such

as changes in macroeconomic variables, forced outages of power generation plants or unexpected contingencies in transmission networks and the like. Such unexpected events often result in price processes following completely different dynamics. Therefore, we need to try other types of affine models to get useful forecasts of the spot prices, and this could be one of the most interesting area of future research.

Bibliography

- [1] S.L. Puller, *Pricing and firm conduct in Californias deregulated electricity market*, University of California Energy Institute, The US Power Market, Risk Publications, Tech. Report, 2002.
- [2] A. Eydeland and G. Hélyette, *Pricing power derivatives*, Risk, 1998.
- [3] J.S. Deng, *Valuation of investment and opportunity-to-invest in power generation assets with spikes in electricity price*, School of Industrial & Systems Engineering, Georgia Institute of Technology, Tech. Report, 2003.
- [4] H. Geman and A. Roncoroni, *A class of marked point processes for modeling electricity prices*, Department of Finance, ESSEC Business school, Tech. Report, 2001.
- [5] D. Duffie, J. Pan, and K.J. Singleton, *Transform analysis and asset pricing for affine jump-diffusion*, *Econometrica* (68), 2000.
- [6] D.S. Bates, *Empirical Option Pricing: A Retrospection*, University of Iowa and the National Bureau of Economic Research, Tech. Report, 2002.
- [7] Q. Dai and K.J. Singleton, *Specification analysis of affine term structure models*, Graduate School of Business Stanford University, Working Paper, 1998.
- [8] G. Chacko and S. Das, *Pricing interest rate derivatives: a general approach*, NBER Working Paper, 2000.

- [9] S. Das and S. Foresi, *Exact solutions for bond and option prices with systematic jump risk*, Review of Derivatives Research, Vol. 1, 1996.
- [10] J.S. Deng, *Pricing electricity derivatives under alternative stochastic spot price models*, University of California Energy Institute, Tech. Report, 1998.
- [11] P. Villaplana, *Pricing power derivatives: a two-factor jump-diffusion approach*, University Carlos III Madrid, Job Market Paper, 2003.
- [12] K.J. Singleton, *Estimation of affine asset pricing models using the empirical characteristic function*, Stanford University and NBER, Journal of Econometrics, 2001.
- [13] G. Chacko and L.M. Viceria, *Spectral GMM estimation of continuous-time processes*, Graduate School of Business Administration, Harvard University, Tech. Report, 2001.
- [14] T. Daniel, J. Doucet and A. Plourde, *Electricity Industry Restructuring: The Alberta Experience*, School of Business, University of Alberta, Working Paper, 2001.
- [15] <http://www.energy.gov.ab.ca/com/Electricity/Introduction/Electricity.htm>
- [16] R. Elliott, G. Sick and M. Stein, *Pricing electricity calls*, University of Alberta and University of Adelaide, University of Calgary, University of Oregon, Tech. Report, 2000.
- [17] G. Lee and B. Jr, *Stochastic Financial Models for Electricity Derivatives*, University of Stanford, Ph.D. Thesis, 1999.

- [18] L.L. Yousef, Derivation and Empirical Testing of Alternative Pricing Models in Alberta's Electricity Market, University of Calgary, Thesis, 2001.
- [19] P. Wilmott, Derivatives, John Wiley and Sons, 1998.
- [20] D. Pilipovich, *Energy Risk: Valuing and Managing Energy Derivatives*, McGraw-Hill, New York, 1997.
- [21] Y. Aït-Sahalia, *Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach*, *Econometrica*, Vol.70, No.1, 2002.
- [22] J.J. Lucia and E.S. Schwartz, *Electricity prices and power derivatives: evidence from the Nordic Power Exchange*, *Review of Derivatives Research* (5), 2002.
- [23] L. Clewlow and C. Strickland, *Energy Derivatives: Pricing and Risk Management, Pricing and Risk*, LACIMA Publications, 2000.
- [24] J.S. Deng, *Stochastic models of energy commodity prices: Mean-reversion with jumps and spikes*, University of California Energy Institute, Tech. Report, 1998.
- [25] H. Zhou, *Jump-Diffusion term structure and Itô conditional moment generator*, Federal Reserve Board, Tech. Report, 2001.
- [26] C.R. Knittel and M.R. Roberts, *An empirical examination of deregulated electricity prices*, University of California Energy Institute, Working Paper, 2001.
- [27] C. Ball and W.N. Torous, *A simplified jump process for common stock returns*, *Journal of Financial and Quantitative Analysis* (18), 1983.

- [28] S. Beekers, *A note on estimating the parameters of the diffusion-jump model of stock returns*, Journal of Financial and Quantitative Analysis (16), 1981.
- [29] M. Carrasco and J.P. Florens, *Estimation of a mixture via the empirical characteristic function*, Tech. Report, 2000.
- [30] P. Honoré, *Pitfalls in estimation jump-diffusion models*, center for analytical finance, University of Aarhus, Working Paper Series No.18, 1998.
- [31] A. Lo, *Maximum likelihood estimation of generalized Ito processes with discretely sampled data*, Econometric Theory (4), 1988.
- [32] N.M. Kiefer, *Discrete parameter variation: Efficient estimation of a switching regression model*, Econometrica, 1978.
- [33] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.
- [34] D.R. Smith, *Essays in Empirical Asset Pricing*, The University of British Columbia, Ph.D. Thesis, 2002.
- [35] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics (31), 1986.
- [36] V. Naik, *Option valuation and hedging strategies with jumps in the volatility of asset returns*, Journal of Finance (48), 1993.
- [37] E. Ghysels, A. Harvey, and E. Renault, *Handbook of Statistics*, vol. 14, 1995.
- [38] N. Shephard, *Statistical aspects of ARCH and stochastic volatility.*, Nuffield College, Oxford, Manuscript, 1995.

- [39] P.E. Protter, *Stochastic Integration and Differential Equations*, Springer-Verlag, 2nd Edition, 2003.
- [40] V. Kaminske, *The challenge of pricing and risk managing electricity derivatives*, The US Power Market, Risk Publication, 1997.
- [41] M. Goto and G.A. Karolyi, *Understanding Electricity Price Volatility Within and Across Markets*, The Central Research Institute of Electric Power Industry in Japan and the Dice Center for Financial Economics, Tech. Report, 2003.
- [42] V. Robert and T. Allen, *Introduction to Mathematical Statistics*, Fifth Edition, Prentice Hall, Englewood Cliffs, New Jersey, 1995.
- [43] G.J. Jiang and J.L. Knight, *Estimation of continuous time processes via the empirical characteristic function*, *Journal of Business and Economic Statistics*, 2001.
- [44] G.R. Grimmett and D.R. Stirzaker, *Probability and Random Processes*, Oxford Science Publications, 2nd Edition, 1982.
- [45] B. Dupire, *Pricing with a smile*, RISK, Vol. 7, 1994.
- [46] T. Coleman, Li and A. Verma, *Reconstructing the unknown volatility function*, *Journal of Computational Finance*, Vol. 2, Number 3, 1999.
- [47] B. Hamida and R. Cont, *Recovering volatility from option prices by evolutionary optimization*, Manuscript, CNRS., 2004.
- [48] R. Cont and P. Tankov, *Financial Modeling with Jump Processes*, Chapman and Hall/CRC, 2004.

- [49] L. Hansen, *Large sample properties of generalized method of moments estimators*, *Econometrica*, Vol. 50, 1982.
- [50] L. Matyas, *Generalized Method of Moments Estimation*, Cambridge, 1999.
- [51] G. Schwartz, *Estimating the dimension of a model*, *Annals of Statistics* (6), 1978.
- [52] S.R. Das, *The surprise element: jumps in interest rates*, Santa Clara University, Tech. Report, 2000.