

Filtering and parameter estimation for electricity markets

by

Alberto Molina-Escobar

B.Sc., Universidad Nacional Autónoma de México, 1996

M. S., Universidad Nacional Autónoma de México, 2000

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Mathematics)

The University Of British Columbia

October, 2009

© Alberto Molina-Escobar 2009

Abstract

The growing complexity of energy markets requires the introduction of increasingly sophisticated tools for the analysis of market structures and for the modeling of the dynamics of spot market and forward prices. In order for market participants to use these markets in an efficient way, it is important to employ good mathematical models of these markets. This has proved to be particularly difficult for electricity, where markets are complex, and exhibit a number of unique features, mainly due to the problems involved in storing electricity.

In this thesis we propose three models for electricity prices. All are multi-factor models, that is, as well as an observable spot price they assume the existence of an unobservable 'long term mean' process. The introduction of such additional processes helps to explain the relation between spot and futures prices. In the first part of the thesis we introduce a two factor Gaussian model for prices. Using the Kalman filter, and based on both spot and forward prices, we successfully estimate parameters for simulated data. We then estimate parameters for the German EEX market, and compare our fitted model with the observed prices. We find that this model does capture some features of the EEX market, but it fails to exhibit the price spikes which are a prominent feature of true spot prices. We therefore introduce a second model, which includes jumps. The inclusion of jumps has the potential to give a better explanation of the behavior of electricity prices, but it creates difficulties in the estimation of parameters. This is because as the model noise is non-Gaussian the Kalman filter cannot be applied satisfactorily. We implement the particle filter adopting the Liu & West approach for the jump model. This method allows us to identify the hidden process in the model, and to estimate a small number of parameters. The third model is a new model for electricity prices based on the inverse Box-Cox transformation. This model is non-linear with Gaussian noise, and can generate price spikes using fewer parameters than a multi-factor jump-diffusion model. In this context, we successfully applied the Unscented Kalman filter to estimate the parameters.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	viii
Dedication	ix
1 Introduction	1
1.1 Commodity markets	1
1.2 Electricity markets	2
1.3 The relationship between spot and futures prices	6
1.4 Previous work	8
2 Filtering	25
2.1 State space formulation	25
2.2 The Kalman filter	27
2.3 The unscented Kalman filter	29
2.4 Particle filter	32
2.5 Parameter estimation via maximum likelihood	38
2.6 Parameter estimation via Bayesian methods	43
3 MROU model	44
3.1 Double mean-reversion model	44
3.2 Radon-Nikodym theorem for Ornstein-Uhlenbeck processes	45
3.3 Future price	49
3.4 Formulation in Kalman filter terms	51

Table of Contents

3.5	Empirical results	56
3.5.1	Simulated data	56
3.5.2	The German electricity market	58
4	MROU with jumps	66
4.1	Description of the model	66
4.2	Valuation of electricity futures	67
4.3	Particle filter setup	70
4.4	Simulated data with known parameters	76
4.5	Likelihood function estimation	77
4.6	Sequential parameters	78
4.7	Empirical results	80
5	NLMROU model	89
5.1	The model	89
5.2	Future price	90
5.3	Unscented Kalman filter setup and estimation procedure	93
5.4	Simulation results	95
5.5	Parameter estimation based on historical data	95
6	Conclusions	101
6.1	Future work	102
	Bibliography	105
	Appendix A	113
	Appendix B	116

List of Tables

1.1	Models for electricity prices.	22
3.1	The data are taken from Hikspoors & Jaimungal (columns 1 and 4), and Nomikos & Soldatos (columns 2 and 3).	47
3.2	Five different maximization runs, on the same set of simulated data.	57
3.3	Estimation using one futures contract (average of 50 simulations).	57
3.4	Estimation using two futures contracts ($n=300$).	58
3.5	Estimated values for $f(t)$ by least-squares fitting.	60
3.6	Estimated values for the EEX market using S_t and $F(t, T_1)$	63
3.7	Estimated values for the EEX market using S_t , $F(t, T_1)$ and $F(t, T_2)$	63
3.8	The table shows the first four moments of the logarithmic deseasonalized price returns of observed data and the average of 50 simulated trajectories.	65
4.1	Sample of 8 estimated values for λ_X	84
4.2	Individual estimates for parameters in MROU with jumps model.	84
5.1	Five different maximization runs with 800 observations.	96
5.2	Estimation using one futures contract ($n = 800$).	97
5.3	Estimated values for the EEX market using S_t , and $F(t, T_1)$	98
5.4	The table shows the first four moments of the logarithmic deseasonalized price returns of observed data and the average of 50 simulated trajectories.	100

List of Figures

1.1	Classification of commodity markets.	1
1.2	World net electricity consumption 2004-2030.	2
1.3	If the total load is low, the plants with the lowest variable production costs are used (nuclear, hydro); if the total load is high, gas or oil fired plants with high fuel cost are running additionally, producing a huge effect on the price.	3
1.4	Seasonal patterns by hours and by week for the German market.	4
1.5	Average daily spot price in German market for years 2002-2007	4
1.6	The factors exerting a major influence on electricity wholesale price.	5
2.1	A graphical representation of the particle filter with importance sampling and resampling.	37
3.1	Sampling distributions with (column 4) where we add 10% deviation from Hikspoors & Jaimungal's parameters (column 1).	48
3.2	The upper graph shows the simulated spot price S_t and the future price $F(t, T_1)$ with maturity of one month. The lower graph is the long-term mean process L_t	55
3.3	Electricity spot and nearby monthly futures price in German market.	60
3.4	The upper graph shows the log spot-price of the EEX market and the seasonal component $\hat{h}(t)$ and the lower graph the deseasonal series $X_t = \log \tilde{S}_t - \hat{h}(t)$	61
3.5	The upper shows the spot price \tilde{S}_t with $\exp\{\hat{h}(t)\}$ and the lower graph the deseasonal spot price $S_t = \exp\{X_t\}$	62
3.6	Simulation of spot and future prices (upper graph) and long-term mean process (lower graph) using estimate values for part 1.	64

List of Figures

4.1	Plot of the true state L_t and estimate of the particle filter. . .	76
4.2	The log-likelihood for different λ_X values.	77
4.3	The graph shows a simulated trajectory of the MROU model with jumps.	82
4.4	The graph shows the upward and downward jumps for the simulated trajectory of the MROU model with jumps.	83
4.5	Sample particle filter trajectories for the estimate of λ_X	84
4.6	Sample particle filter trajectories for the estimate of η_u and η_d using Liu and West approach.	87
4.7	Sample particle filter trajectories for the estimate of σ_X and λ_L using Liu and West approach.	88
5.1	The graph shows a simulated trajectory of the NLMROU model.	96
5.2	Plot of the true and estimated processes L_t and X_t of the NLMROU model (first 80 observations).	97
5.3	Simulation of spot and future prices price using estimated val- ues for the whole data set.	100

Acknowledgements

First and foremost I must acknowledge my thesis supervisor, Martin Barlow. Throughout my time as a graduate student Martin has provided me with unending guidance in research, constant encouragement and many years of financial support. Without his undying support this thesis would not have been possible.

I am also grateful for the helpful comments of Rachel Kuske, Ulrich Horst, Joel Friedman and Kevin Murphy. Many thanks also to Arnaud Doucet for his feedback and recommendations that help me improve my project. I would also like to thank Gabriel Mititica for his help and encouragement.

Thank you also to CONACyT for its financial support during the first years of my program.

To my wife and daughter.

Chapter 1

Introduction

In this thesis we are interested in the commodity futures markets, and in particular in electricity futures markets.

1.1 Commodity markets

In this section we are going to describe some of the unique characteristics present only in commodities markets. Figure 1.1 shows the three fundamentals groups of commodities.

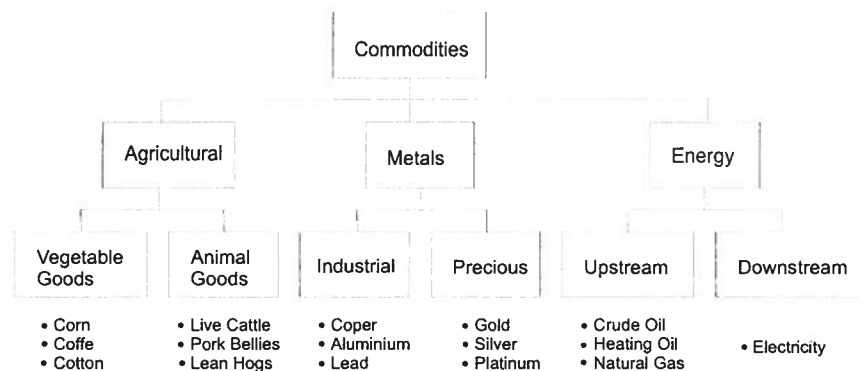


Figure 1.1: Classification of commodity markets.

Commodities markets exhibit some characteristics that are not present in financial markets due to physical constraints and also due to the variation of demand due to changes in consumption. The commodity spot price is defined by the intersection of supply and demand curves. Thus the spot price can be affected by changes in consumption, production or inventory. Unlike financial assets, which are traded for investment purposes, commodities are traded in order to be consumed or used in an industrial process, with the partial

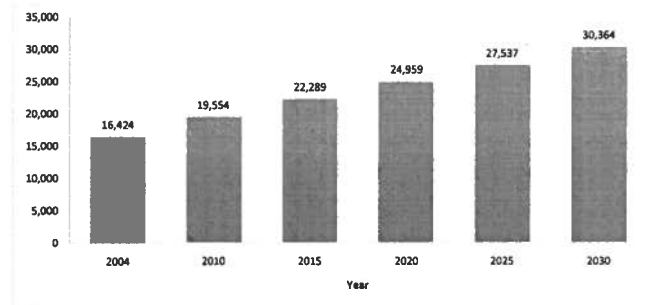


Figure 1.2: World net electricity consumption 2004-2030.

exception of some precious metals. This close link with the real economy causes commodities prices to have seasonal behavior and also mean-reversion [31]. This is one reason why many of the standard financial theories may not be applicable to commodities markets.

1.2 Electricity markets

Among the various commodities the energy market is the most recent market to be transformed. Since the early 1990s, electricity markets have been and continue to be developed as a result of the deregulation of electricity markets worldwide. In many regions the market structure has moved from a monopolistic to a competitive one. Traditionally, there was only one company or government agency that produced, moved, distributed, and sold electricity power and services. This transformation has been already taken place in the Americas (parts of Canada and US, Argentina, Chile, Peru, Paraguay and Colombia) in Europe (Norway, Finland, Denmark, Germany, France, Netherlands, Spain, Poland, and Romania) and in the Asia/Pacific region (Australia, New Zealand and Japan).

In theory, deregulating the electricity market should increase the efficiency of the industry by producing electricity at lower costs and passing those cost savings on to customers [39]. Electricity is a growing market. In 1973 electricity consumption accounted for 11% of the total world energy

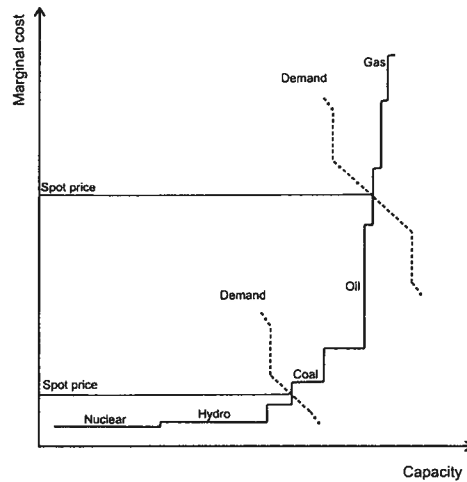


Figure 1.3: If the total load is low, the plants with the lowest variable production costs are used (nuclear, hydro); if the total load is high, gas or oil fired plants with high fuel cost are running additionally, producing a huge effect on the price.

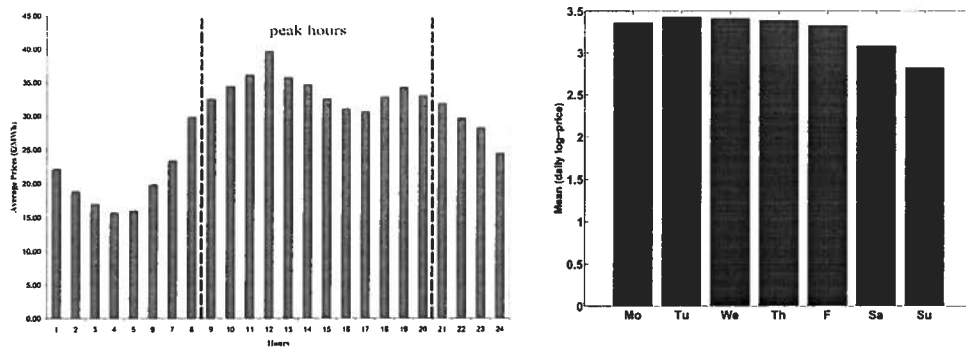
demand and has grown to 18% today. The absolute growth rate of electricity consumption in the future is estimated at an average of 2.4% per year. The projected growth in electricity consumption is shown in Figure 1.2.¹

Electricity is considered a secondary energy source, which means it is created from the conversion of other sources of energy, such as coal, oil, natural gas, nuclear power, or hydropower, all of which are referred to as primary energy sources. To understand the behavior of electricity prices we have to note that electricity possesses a unique feature; it is very difficult and expensive to store and quite difficult to transmit from one region to another.

As a result of this, the spot price of electricity is set by the short-term supply-demand equilibrium, and supply and demand must be in balance at each time. Figure 1.3 displays a schematic supply-demand curve. The supply and demand are affected by many factors that influence the seasonality and volatility of prices. For example, supply may be affected for transition constraints (breakdowns) or fluctuation of fuel prices (oil, gas). Demand exhibits seasonal fluctuations, which are due to climate conditions. In ad-

¹Source: (EIA) System for the Analysis of Global Energy Markets (2007).

Chapter 1. Introduction



(a) Hours average price from Jan-Dec 2002. (b) Daily average log-price throughout the week from 12/2002 to 05/2005.

Figure 1.4: Seasonal patterns by hours and by week for the German market.

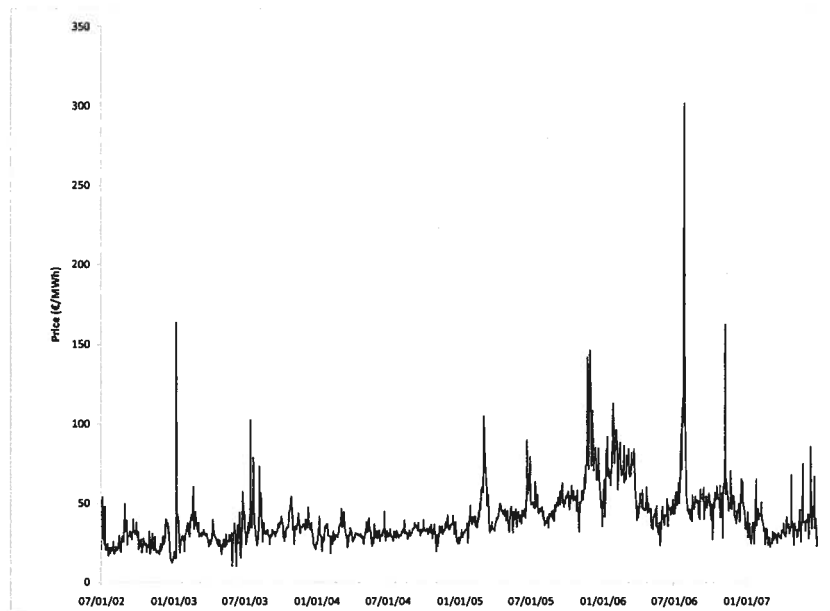


Figure 1.5: Average daily spot price in German market for years 2002-2007

dition, electricity demand is also not uniform through the week. It peaks during weekday working hours and is low during nights, holidays and weekends due to lower industrial activity, see Figure 1.4. Also unexpected weather conditions can cause abrupt and dramatic disruptions, producing jumps and spikes in the spot price. Finally the constraints on transmission mean that power markets are geographically distinct. In some markets (such as Alberta or Norway), demand is higher in the winter months due to the use of power for heating. In other markets, such as California power, usage peaks in the summer due to use of electricity for air conditioners. Figure 1.6 shows the factors that influence the determination of electricity prices.²

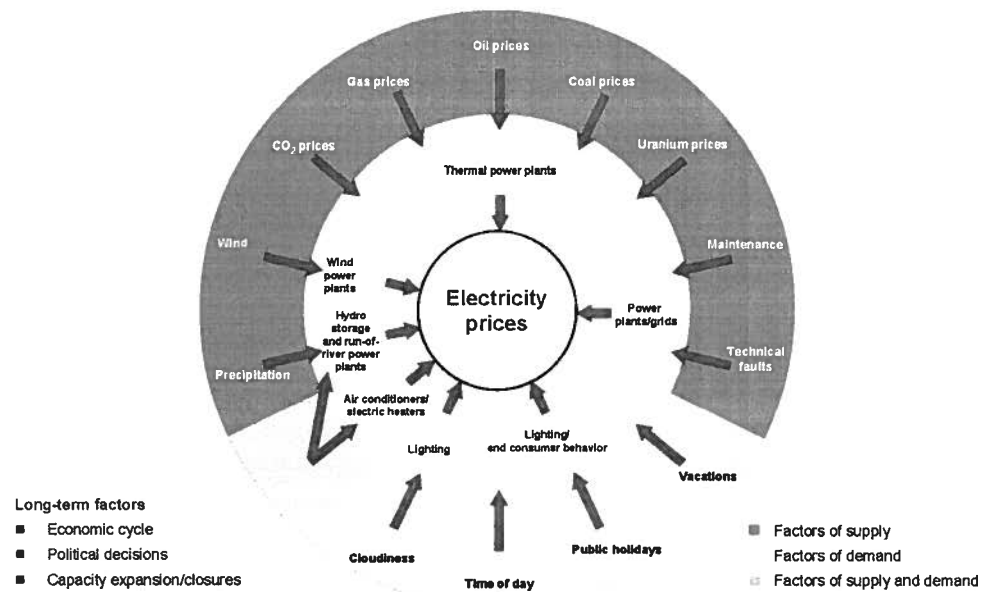


Figure 1.6: The factors exerting a major influence on electricity wholesale price.

The most unusual feature of electricity spot prices is the presence of “price spikes”; a phenomenon which does not have any parallel in other commodity markets. See for example Figure 1.5 which shows that in some days in July

²Source: (RWE AG) Shares of Primary Energy Sources in Total Electricity Generation in Europe (2008).

2006 the spot price in Germany reached 300 €/MWh, compared with a normal daily price of 30-50 €/MWh.

If such an event occurred for a conventional commodity, such as say copper, holders of the material would be able to made substantial profits by selling the commodity during the spike, and then repurchasing it at a normal price a few days later. But, because, it cannot easily be stored, this is not possible with electricity.

1.3 The relationship between spot and futures prices

The relationship between the spot price and the futures (forward) prices is important for risk management and option pricing theory [30]. Across all these commodities ranging from agricultural products to pure financial assets certain common principles of futures valuation and futures price behavior apply.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space endowed with the natural filtration $\{\mathcal{F}_t\}_{t \geq 0}$. In the financial world, the relation between spot and future prices, under the risk-neutral measure \mathbb{Q} , is given by the formula

$$F(t, T) = \mathbb{E}_t^{\mathbb{Q}} \left(S_T \exp \left\{ \int_t^T r_u du \right\} \right) \quad (1.1)$$

where r_u is the risk-free interest rate. The proof is based on the no-arbitrage argument (see [9]), which proceeds by comparing returns on a portfolio consisting of the future contract with one consisting of cash and the commodity.

However, unlike financial assets, storage of commodities is costly. Consequently, physical ownership of the commodity carries an associated flow of services. On the one hand, the owner enjoys the benefit of direct access, which is important if the commodity is to be consumed. On the other hand, postponing consumption and storing the commodity means that storage expenses have to be paid. The net flow of these services per unit of time is called the convenience yield c_t . Since the convenience yield is the result of subtracting the cost of storage from the benefit attached to the physical commodity it can be both positive or negative at different times. (A positive convenience yield implies an instantaneous benefit from holding the commodity, a negative one an instantaneous cost.)

Again, by a no-arbitrage argument, the relationship between the spot and forward price is given by:

$$F(t, T) = \mathbb{E}_t^{\mathbb{Q}} \left(S_T \exp \left\{ \int_t^T (r_u - c_u) du \right\} \right), \quad (1.2)$$

where c_u is the instantaneous forward convenience yield [31]. Note that the convenience yield plays the same role as dividends play for stocks.

Some authors have argued that as a consequence of the non-storability of electricity the notion of convenience yield is irrelevant in power markets. Therefore the relation between spot and futures (forward) prices cannot be established through the no-arbitrage argument (see [14, 32, 33]). For example, Geman and Roncoroni comment in [32]:

“Our view is that a convenience yield does not really make sense in the context of electricity: since there is no available technique to store power (outside of hydro), there cannot be a benefit from holding the commodity, nor a storage cost. Hence, the spot price process should contain by itself most of the fundamental properties of power.”

Other financial theories view the futures (forward) prices $F(t, T)$ and the expected future spot price $\mathbb{E}_t(S_T)$ as related but not identical. The difference is the risk premium, i.e.

$$F(t, T) = \mathbb{E}_t(S_T) + \pi(t, T). \quad (1.3)$$

The full specification is not straightforward to establish. The theory of a positive risk premium is termed normal backwardation. The opposite situation where the futures prices is set above the expected future spot price (a negative risk premium) is called contango.

An alternative approach is the actuarial one, which values a forward contract as its discounted expected real world payoff, see [44]. This is the approach we will adopt in this thesis: we will assume that the relation between spot and futures prices is given by (1.2), and that the risk free interest rate r_u and convenience yield c_u are constant.

1.4 Previous work

The main motivation for the development of models for electricity prices is the need for such models by market participants. For example, a power company has the choice of selling its power either on the spot or forward market, and would wish to make the optimum choice. In addition there is the need to price derivatives such as forwards, options and swaps. Hence the model should be sufficiently sophisticated for realistic modelling but sufficiently simple for pricing of derivatives. This issue is very important for computing risk measures, testing hedging strategies and evaluating investment policies.

Various approaches have been developed to describe the stochastic price process in energy markets. There are significant parallels between commodity markets and interest rate markets. For commodity markets, the traded assets are both the spot and various forward or future contracts. For interest rates, the main traded assets are futures (represented by different types of bonds), while the spot or instantaneous rate of interest plays a more minor role. Given these parallels, it is natural to use the interest rate theory as a base for electricity price models.

In general, interest rate models can be separated into two categories: short-rate model and forward-rate models. The short-rate models describe the evolution of the instantaneous interest rate as stochastic process, and the forward-rate models capture the dynamics of the whole forward curve (Heath-Jarrow-Morton models). These interest rate models are then applied to arrive to arbitrage-free pricing of bonds or other derivative products.

The same division of models arises for power prices, where the models may be broadly divided into two groups:³ statistical models (spot price based models) and fundamental models (forward based models).

For the forward based models, the futures prices are the main objects of study, and the dynamics of the whole futures prices curve is modeled using the Heath-Jarrow-Morton [42] theory for interest rates. See for example, Clewlow and Strickland (1999) [19] and Manoliu and Tompaidis (2002) [61], and for more recent papers see Borovkova (2006) [10], Koekebakker and Ollmar (2005) [55].

³There is another approach based on econometric time series model that we will not consider in this work (see [56, 63])

A general discussion of HJM-type models in the context of power futures is given in Benth and Koekebakker (2008) [8]. They dedicate a large part of their analysis to the relation of spot, forward and swap-price dynamics and derive no-arbitrage conditions in power future markets and conduct a statistical study comparing a one-factor model with several volatility specifications using data from the Nord Pool market.

The disadvantage of such approaches is that futures prices do not reveal information about price behavior on a daily timescale and provide a poor approximation to the complex observed spot behavior in power markets.

In this thesis, following most of the literature, and the philosophy outlined by Geman and Roncoroni above, we will consider spot based models. In principle these models should provide a reliable description of the evolution of electricity prices. Moreover, these models are versatile in the sense that it is relatively simple to aggregate characteristics to an existing family or class of models by for example adding a seasonality function.

Securities (stocks) are usually modeled by Geometric Brownian Motion with drift

$$S_t = S_0 \exp \{at + \sigma W_t\},$$

as in the famous Black-Scholes model.

This model is not found suitable for commodities, since ‘mean reversion’ is typical feature of these markets [30, 31, 66].

The simplest stochastic process with mean-reverting behavior is the Ornstein-Uhlenbeck process [66]. Here the process X_t is a diffusion process satisfying the stochastic differential equation

$$dX_t = -\lambda(X_t - a)dt + \sigma dW_t \quad (1.4)$$

where W_t is a standard Brownian motion, σ the volatility of the process, and λ the velocity with which the process reverts to its long term mean a . Many electricity price models use this process or variants as a basic building block.

For example, Lucia and Schwartz (2002) [59], give models of the form

$$S_t = h(t) + X_t \quad (1.5)$$

or

$$S_t = \exp \{h(t) + X_t\} \quad (1.6)$$

where S_t is the spot price, X_t is an Ornstein-Uhlenbeck process, and $h(t)$ is a deterministic component, intended to account for seasonal and weekly effects. Benth et al. (2008) [6] called models like (1.5) ‘arithmetic models’ and (1.6) ‘geometric models’, i.e. geometric models represent the logarithmic prices by a sum of processes.

The incorporation of a deterministic component of this kind is an important feature of nearly all spot price models. Spot prices are higher on weekdays than on weekends, due to higher demand, so a correction $h(t)$ which compensates for this is essential. See for example Figure 1.4b.

Spot price models can be divided into ‘single’ factor or multi-factor models. For single factor models the spot price is itself a Markov process, while in multi-factor models the spot price is a function $S_t = g(X_t^1, \dots, X_t^k)$ of a multidimensional Markov process. Here $g : \mathbb{R}^k \rightarrow \mathbb{R}^+$, and as g is not one-to-one these models have unknown or hidden components.

As well as the model of Lucia and Schwartz mentioned above, other single factor models are in Cartea and Figueroa (2005) [16], Barlow (2002) [3], Kanamura and Ohashi (2007) [49], and Geman and Roncoroni (2006) [32].

Many of these models, unlike that of Lucia and Schwartz, include mechanisms to take account of price spikes. One of the simplest of these is in Cartea and Figueroa [16], which adds a jump term to the Ornstein-Uhlenbeck process:

$$\begin{aligned}\log S_t &= h(t) + Y_t, \\ dY_t &= -\alpha Y_t dt + \sigma dW_t + J_t dN_t,\end{aligned}\tag{1.7}$$

where W_t is a Brownian motion, $h(t)$ is assumed to capture the seasonal patterns of the spot price, and the third term $J_t dN_t$ enables the process to have discrete random spikes: these are a combination of a Poisson process, which determines the jump frequency, and a jump-size distribution, which gives the jump magnitude conditional on a jump occurring.

In (1.7) the process dN_t is approximated by a Bernoulli process with parameter ldt and J_t is log-Normal, i.e. $\log J \sim N(-\sigma^2/2, \sigma^2)$. Cartea and Figueroa apply this one-factor mean-reverting jump diffusion model for the electricity spot price, adjusted to incorporate seasonality effects and derive the corresponding forward in closed-form to the England and Wales market.

However, the rather short period for which electricity prices were available and the small number of spikes caused difficulties with parameter estimation.

Such models require a high speed of mean reversion in order to reduce the spot price following a large positive jump, and this has the effect of removing too much variability in the series over the non-jump time-periods.

Barlow [3] introduces a nonlinear Ornstein-Uhlenbeck model for spot power prices.

The price is obtained by matching the demand level with a deterministic supply function which must be nonlinear to account for price spikes. He proposes the inverse function of the Box-Cox transformation.

$$S_t = \begin{cases} f_\alpha(X_t), & 1 + \alpha X_t > \varepsilon_0 \\ \varepsilon_0^{1/\alpha}, & 1 + \alpha X_t \leq \varepsilon_0 \end{cases}$$

$$dX_t = -\lambda(X_t - a)dt + \sigma dW_t,$$

where

$$f_\alpha(x) = (1 + \alpha x)^{1/\alpha}, \quad \alpha \neq 0 \quad f_0(x) = e^x.$$

When $\alpha = 0$, an exponential Ornstein-Uhlenbeck process is retrieved for S_t . The case $\alpha = 1$ yields a regular Ornstein-Uhlenbeck process. The model has been estimated by maximum likelihood on the Alberta and California markets.

Another paper sharing the same theoretical idea is found in Kanamura and Ohashi [49]. Instead of using the inverse function of the Box-Cox transformation they assume that the supply curve has a ‘hockey stick’ shape. Setting $X_t = D_t - \overline{D_t}$, $\overline{D_t}$ describes the seasonal component and

$$S_t = \begin{cases} (a_1 + b_1 D_t), & D_t \leq D_0 \\ (a_2 + b_2 D_t), & D_t > D_0 \end{cases}$$

$$dX_t = (\mu - \lambda X_t)dt + \sigma dW_t.$$

This model captures the observed mean-reverting behavior of electricity markets and it accounts very well for the observed price spikes, allowing for a better fit to market data. But the assumption of a deterministic supply

function is probably too restrictive since this implies that spikes can only be produced by surges in demand.

Geman and Roncoroni [32] built up a jump-reversion model for electricity spot prices. The model assumes that the natural logarithm of power price dynamics is described by a stochastic differential equation

$$dE(t) = [h(t) + \theta(\mu(t) - E(t^-))]dt + \sigma dW(t) + f(E(t^-))dJ(t), \quad (1.8)$$

where $h(t)$ is a deterministic seasonality function, θ is the mean reversion speed, and σ is a constant instantaneous volatility, $\mu(t)$ is the mean reversion level. The process reverts to a deterministic mean level rather than the stochastic pre-spike value. The last term in equation (1.8) represents the discontinuous part of the model featuring price spikes. This effect is characterized by three quantities defining occurrence, direction, and size of jumps. f is a function which is ± 1 depending on the level of the spot price.

$$f(E(t)) = \begin{cases} +1, & \text{if } E(t) < \tau(t) \\ -1, & \text{if } E(t) \geq \tau(t) \end{cases}$$

The process $J(t)$ is a time-inhomogeneous compound Poisson process with intensity function

$$\lambda(t) = \kappa \left(\frac{2}{1 + |\sin[\pi(t - \gamma)/\delta]|} - 1 \right)$$

where the expected maximum number of jumps per year is represented by κ . Jump sizes are modeled by a sequence of independent and identically distributed truncated exponential variables.

This model generates trajectories similar to those observed in the electricity market, and also it gives a good fit of the empirical moments of order 1, 2 and 4, i.e. mean, variance and kurtosis.⁴

Neither of the last two models includes the convenience yield as a factor, nor considers the valuation of futures contracts or any other kind of derivative. The single factor models are quite tractable and their parameters are

⁴The kurtosis of a random variable X with mean m and variance σ^2 is defined by: $\kappa = \frac{\mathbb{E}((X-m)^4)}{\sigma^4}$. When κ is much greater than 3, it means that the density in the tail is higher than that which prevails for a Gaussian distribution.

relatively easy to estimate. However they have a serious limitation: they do a poor job explaining the relation between spot and futures prices, see [3] and [12]. This limitation can be avoided if changes in spot prices are allowed to depend on more than one factor.

The copper mine example of Brennan and Schwartz (1985) [11] assumed that the spot price followed a geometric Brownian motion and incorporated a convenience yield to their model, assuming it was proportional to the spot price.

$$\begin{aligned}dS_t &= \mu S_t dt + \sigma S_t dz, \\C(S, t) &= cS.\end{aligned}$$

The idea of a constant convenience yield only holds under restrictive assumptions, since the theory of storage is rooted in an inverse relationship between the convenience yield and the level of inventories. Gibson and Schwartz (1990) [35] take an important step to a more realistic model of the economy by introducing a stochastic convenience yield rate. The spot price S_t of the commodity is described by a geometrical Brownian motion and the convenience yield rate δ_t is described by an Ornstein-Uhlenbeck process with equilibrium level α and rate of mean-reversion κ :

$$\begin{aligned}dS_t &= (\mu - \delta_t) S_t dt + \sigma_1 S_t dz_1, \\d\delta_t &= \kappa(\alpha - \delta_t) dt + \sigma_2 dz_2, \\dz_1 dz_2 &= \rho dt.\end{aligned}$$

Significant contributions have been made by Schwartz (1997) [73]. He reviewed one and two factor models and developed a three factor model under stochastic convenience yield and interest rates. Including the interest

as a third factor makes forward and futures prices different.

$$\begin{aligned} dS_t &= (r_t - \delta_t)S_t dt + \sigma_1 S_t dz_1, \\ d\delta_t &= \kappa(\alpha - \delta_t)dt + \sigma_2 dz_2, \\ dr_t &= a(m - r_t)dt + \sigma_3 dz_3, \\ dz_1 dz_2 &= \rho_1 dt, \quad dz_2 dz_3 = \rho_2 dt, \quad dz_1 dz_3 = \rho_3 dt. \end{aligned}$$

This model was originally developed for two commercial commodities (copper and oil). He used the Kalman filter algorithm to estimate the parameters in the models.

In the paper by Lucia and Schwartz [59], where they analyze the Nordic Power market, the spot price is modeled by

$$\begin{aligned} S_t &= h(t) + X_t + Y_t, \\ dX_t &= -\lambda X_t dt + \sigma_X dW_X, \\ dY_t &= \mu dt + \sigma_Y dW_Y, \\ dW_X dW_Y &= \rho dt. \end{aligned}$$

The function $h(t)$ is deterministic, and it is intended to capture the predictable component in the spot price, i.e. seasonal effects. This function distinguishes between weekdays, and includes a monthly seasonal component employing dummy variables. The idea of this model is to have a non-stationary process for the long-term equilibrium price level Y and short-term mean-reverting component X . They estimated all the parameters simultaneously by nonlinear least squares methods.

The multi-factor models described so far do not capture one of the most characteristic feature of power prices, jumps or spikes. Several authors, Deng (2000) [24], Villaplana (2003) [80], and Xiong (2004) [82] extend such models to even more factors with both diffusion and jumps. In the work of Villaplana power prices are modeled according to non-observable state variables that account for the short-term movements and long-term trends in electricity

prices.

$$\ln S_t = h(t) + X_t + Y_t$$

$$dX_t = -\kappa_X X_t dt + \sigma_X dW_1 + J_u dN(\lambda_u) - J_d dN(\lambda_d)$$

$$dY_t = -\kappa_Y (\mu - Y_t) dt + \sigma_Y dW_2$$

$$dW_1 dW_2 = \rho dt.$$

The jump components are characterized by $N(\lambda_u)$, and $N(\lambda_d)$, i.e. Poisson processes with intensities λ_u and λ_d respectively, and by random jumps of size J_u and J_d with some specified distribution (Gaussian/Exponential).

Deng (2000) and Villaplana (2003) set their models in the affine jump diffusion (AJD) framework which enables them to use transform results of Duffie et al. (2000) [27] to derive tractable closed-form solutions for a variety of contracts. Deng proposes more sophisticated mean-reverting jump diffusion models with deterministic/stochastic volatility and regime switching, which may be a good way of addressing the dramatic changes in spot prices. However the trajectories produced by the model are fairly different from the ones observed in the market.

Cartea and Villaplana (2008) [17] build a model for wholesale power prices explained by two state variables (demand and capacity) and calculate the forward premium. Writing D_t , C_t for the demand and capacity, they model D_t and C_t by

$$D_t = f_D(t) + X_t^D,$$

$$C_t = f_C(t) + X_t^C,$$

where f_D , f_C are deterministic functions, and X_t^D , X_t^C are independent Ornstein-Uhlenbeck processes. They then take the spot price as given by

$$S_t = \beta \exp \{ \alpha D_t + \gamma C_t \}.$$

They perform empirical research embracing PJM, England and Wales, and Nord Pool markets. They find that, depending on the market and the period under study, the volatility of capacity and the market price of capacity risk could either put upward or downward pressure on forward prices. They also

find that the forward premium follows a seasonal pattern, being positive in the months of high volatility of demand and close to zero or even negative in the months of low volatility of demand.

Inspired by Cartea and Villaplana (2008), Lyle and Elliott (2009) [60] present an hybrid model that uses a supply-demand approach for price electricity derivatives. They assume that the system demand $D(t)$ is given by

$$D(t) = f(t) + \hat{D}(t)$$

where $\hat{D}(t)$ is an Ornstein-Uhlenbeck process and $f(t)$ a deterministic function. For the supply side, they model the curve $S(t, P)$ which gives the supply at time t if the price is P . They consider curves of the form

$$S(t, P) = aS_b(t) + b \log(cP + \xi)$$

where $S_b(t)$ is the base portion of the system supply and a , b , c , and ξ are positive constants. They consider two different models for $S_b(t)$: a mean-reversing model, and a Markov chain model. The equilibrium price is given by

$$P(t) = \frac{1}{c} \left(\exp \left\{ -\frac{aS_b(t) - D(t)}{b} \right\} \right).$$

Using these equations Lyle and Elliott are able to obtain closed-form solutions for European options. They test the model on Alberta prices data calculating the first four empirical moments. The model gives a good fit for the mean and standard deviation but not for the skewness and kurtosis.

Benth et al. (2007) [7] propose a non-Gaussian Ornstein-Uhlenbeck process which takes into account seasonality and price spikes. Their model is

$$S(t) = h(t) + X(t)$$

where $S(t)$ is the spot price, $h(t)$ is a deterministic periodic function and $X(t)$ is a sum of independent Lévy-driven Ornstein-Uhlenbeck components.

$$X(t) = \sum_{i=1}^n Y_i(t),$$

$$dY_i(t) = -\alpha_i Y_i(t)dt + \sigma_i(t)dL_i(t), \quad Y_i(0) = y_i.$$

Here the processes $L_i(t)$ are independent, possibly time inhomogeneous pure jump Lévy processes with $\mathbb{E}(L_i^2(1)) < \infty$. $L_i(t)$ can be written in terms of their jump measures $N_i(dt, dz)$, $i = 1, \dots, n$.

$$L_i(t) = \int_0^t \int_0^\infty z N_i(ds, dz).$$

The deterministic predictable compensator of $N_i(ds, dz)$ which is the jump measure of $L_i(t)$ is of the form:

$$v_i(dt, dz) = \rho_i(t) dt v_i(t) dz.$$

Here $\rho_i(t)$ is a deterministic function that controls seasonal variation of the jump intensity, $\sigma_i(t)$ controls the seasonal variation of the jump sizes, α_i is the level of mean reversion, and $L_i(t)$ controls the variation of price such as the daily volatile variation and price spikes. Benth et al. (2007) provide closed-form and semi-closed form solutions for forwards and options on forwards. This model, coupled with a good description of price seasonality, provides a precise characterization of electricity spot price behavior. In addition, due to its arithmetic structure, it is analytically tractable when it comes to futures and other derivatives pricing. Although the model seems to capture the stylized facts of spot prices market such as mean reversion, seasonality and price spikes, the authors did not make a precise statistical analysis of the quality of the model. However, they suggested the particle filter as a possible method to estimate the parameters in the model. Parameter estimation for this model would appear to be a significant challenge.

Hikspoors and Jaimungal (2007) [44] consider two models for oil prices. The first is a two factor version of the model of Schwartz:

$$\begin{aligned} S_t &= \exp\{h(t) + X_t\}, \\ dX_t &= \lambda_X(Y_t - X_t)dt + \sigma_X dW_t, \\ dY_t &= \lambda_Y(\phi - Y_t)dt + \sigma_Y dZ_t, \\ dW_t dZ_t &= \rho dt. \end{aligned} \tag{1.9}$$

(In fact, in order to value spread options, they consider two different commodities satisfying (1.9)). The second is a modification of (1.9) with an

additional jump component to handle price spikes:

$$S_t = \exp\{h(t) + X_t + J_t\}, \quad (1.10)$$

where X_t is given by (1.9), and

$$dJ_t = -\kappa J_{t-} dt + dQ_t \quad (1.11)$$

with Q_t a compound Poisson process, and t^- denotes the instant immediately before time t . Through measure changes induced by a pseudo-numeraire, they obtained, for both models, exchange option and futures prices in closed form both under real-world and risk-neutral measures. Also they consider the problem of model calibration. For the jump model (1.10) they suggest a modification of the procedure of [20]. This is to identify the price jump by searching for days with a price change more than 3 times the standard deviation of the daily price change. (This procedure is then run several times to produce a 'despiked' series).

They do not apply this method to electricity prices, but do estimate parameters for their first model for oil, from a period of about 3 years of data. They consider parameters built under the real world probability \mathbb{P} , and a risk neutral probability \mathbb{Q} . Let us write a bar on the model parameters to denote parameters with respect to \mathbb{Q} . If $\mathbb{P} \sim \mathbb{Q}$ on the filtration generated by both X_t and Y_t , then one has

$$\sigma_X = \overline{\sigma_X}, \quad \sigma_Y = \overline{\sigma_Y}, \quad \rho = \overline{\rho}.$$

The parameter estimation proceeded in a number of steps:

1. Using least squares, the risk neutral parameters were estimated from the forward data, as is the unknown process Y_t .
2. Given (X_t, Y_t) , the remaining real world parameters $\lambda_X, \lambda_Y, \phi$ were estimated by regression.

One surprising feature of the estimates is that they obtain $\rho = -0.96$, so that the long-term process Y_t adds little extra randomness.

The advantage of two or more factor models is that they allow for a good mathematical description of the problem. Furthermore they have a better fit to historical data and provide a better relation between spot and futures prices, see [73].

Many of the models in the literature incorporate ‘regime switching’ components. For example, the basic model of Nomikos and Soldatos (2008) [64] for Nord Pool prices, is similar to that of Hikspoors and Jaimungal:

$$S_t = h(t) + \exp \{X_t + Y_t\}, \quad (1.12)$$

where X_t is an Ornstein-Uhlenbeck type process, and Y_t is a jump process intended to take account of price spikes. However, both X_t and Y_t incorporate regime dependent terms. X_t is given by

$$dX_t = \kappa_1(\varepsilon(R_t) - X_t)dt + \sigma_X dW_t, \quad (1.13)$$

where R_t is the water reservoir level, and is assumed to follow a two-state Markov chain with state space {wet, dry}. (Much of the Nord Pool electricity production is by hydro.) Y_t is driven by a ‘jump process’, and satisfies

$$dY_t = \kappa_2 Y_t dt + dL_t. \quad (1.14)$$

Here L_t is a jump process, with rates and jump distribution dependent on the season.

Davison et al. (2002) [22] they propose a hybrid model based on the ratio $\alpha(t)$ between demand and capacity. At each time step t the spot price S_t is drawn from a distribution $P(t)$ which is a mixture of Gaussian distributions given by

$$P(t) = (1 - \epsilon(\alpha(t)))P_L(t) + \epsilon(\alpha(t))P_H(t).$$

Here $P_H(t)$ is the price-spike distribution, $P_L(t)$ is the low-price distribution, and ϵ is a function of α that represents a relative demand-capacity ratio. The distribution ϵ plays the role of a switching variable that determines whether the price is to be drawn from $P_L(t)$ or $P_H(t)$, i.e. the probability of a spike. They assumed

$$\epsilon(t) = \frac{1}{2} \tanh(20 * (\alpha(t) - 0.85)) + \frac{1}{2}$$

where the constants are determined by historical PJM spot prices. The distributions P_L and P_H are taken to be Gaussian and the function $\alpha(t)$ is deterministic.

It appears that each time step t independent samples are taken from the distribution P_L, P_H . The choice of independent samples from P_L would

lead to highly oscillation prices, and in Anderson and Davison (2008) [1] they replace P_L by a Brownian motion. To test the model, they simulated trajectories and compare statistical moments. Applying the Kolmogorov-Smirnov test, they concluded that their model is able to simulate data that are from a similar distribution to the observed prices.

Erlwein et al. (2008) [29] develop and analyze an exponential Ornstein-Uhlenbeck process with an added jump process based on hidden Markov model (HMM) setting. The jump component is a Poisson process where the mean and variance are controlled by a discrete time HMM. That is, the spot price, that is partially observed (the underlying economic state is hidden) is given by

$$S_t = D_t \exp\{X_t\}$$

where D_t is a deterministic function and

$$dX_t = \alpha(z_t)(\beta(z_t) - X_t)dt + \sigma(z_t)dW_t + J_i dN_t.$$

Here z_t is a Markov chain with 2 or 3 states, and the jump sizes J_i are conditional Gaussian distributed, i.e.

$$J_i|Z_t \sim N(\mu(z_t), \sigma^2(z_t)).$$

They apply the EM algorithm to estimate the parameters in the model using data from the Nord Pool market. The model captures some of the spikes presents in the real data for the 2 and 3-state Markov chain. A puzzling feature of this paper is that the estimates for the transition matrix of z_t suggest that z_t are close to i.i.d. random variables.

In the same theoretical framework a continuous-time process is derived by Kholodnyi (2001) [52], where self-reversing non-Markovian spikes are added to a Markovian regular price process.

One sees in the literature the need to balance two competing demands. Simple models, particularly simple one factor models, have relatively few parameters, and these parameters may be relatively easy to estimate. However, these models generally fail to capture one or more features of real markets of which the most difficult are the existence of price spikes, and the relation between spot and future prices.

The need for a better fit with data leads to more complicated models, usually with hidden variables, and sometimes with multiple regimes. However introducing more factors requires introducing more parameters into the model. Parameter estimation then becomes a significant challenge. Benth et al. (2008) [6] remark:

“The question of estimating such models on data is not an easy one... For multi-factor models this may be an even more challenging problem, involving highly sophisticated estimation techniques”.

Such parameter estimation is, of course, an essential preliminary to the valuation of options or derivatives based on the commodity. Table 1.1 summarize some of the statistical models.

Although the standard statistical procedure for estimation of a partly unobserved process involves filtering methods, few of the papers in the literature use those techniques. An exception is Culot et al. (2006) [21]. They consider a model of the form

$$\log S_t = h(t) + \tilde{\gamma}^T \tilde{X}_t + \gamma^T X_t,$$

where $h(t)$ is deterministic, and \tilde{X}_t , and X_t are spikes and long-term factors. The spike process \tilde{X}_t is an m state Markov regime switching process, while X_t is a 3-dimensional Ornstein-Uhlenbeck type process given by $X_t = (X_t^{(1)}, X_t^{(2)}, X_t^{(3)})$, where

$$dX_t^{(i)} = -\kappa_i dt + \sigma_i dW_t^{(i)},$$

and $W_t^{(i)}$ are independent Brownian motions. After estimating the jump term \tilde{X}_t and subtracting this from the series, the authors combine Kalman filtering techniques with maximum likelihood estimation, using both spot and forward prices, to estimate the parameters for X_t .

Another exception is Kellerhals (2001) [51]. He developed a model for short-term electricity forwards. The suggested stochastic volatility model uses the non-tradeable spot price S_t of electricity and its variance rate v_t as state variables. The stochastic specification of the processes is given by

$$dS_t = \mu S_t dt + S_t \sqrt{v_t} dW_t,$$

$$dv_t = \mu_v dt + \sigma \sqrt{v_t} dZ_t,$$

$$dW_t dZ_t = \rho dt.$$

SPOT PRICE BASED MODELS	
Authors	Model Specification
Gibson / Schwartz (1990)	$dS_t = (\mu - \delta_t)S_t dt + \sigma_1 S_t dz_1$ $d\delta_t = \kappa(\alpha - \delta_t)dt + \sigma_2 dz_2$ $dz_1 dz_2 = \rho dt$
Schwartz (1997)	$dS_t = (r_t - \delta_t)S_t dt + \sigma_1 S_t dz_1$ $d\delta_t = \kappa(\alpha - \delta_t)dt + \sigma_2 dz_2$ $dr_t = a(m - r_t)dt + \sigma_3 dz_3$ $dz_1 dz_2 = \rho_1 dt, \quad dz_2 dz_3 = \rho_2 dt, \quad dz_1 dz_3 = \rho_3 dt$
Lucia / Schwartz (2002)	$S_t = h(t) + X_t + Y_t$ $dX_t = -\lambda X_t dt + \sigma_X dW_X$ $dY_t = \mu dt + \sigma_Y dW_Y$ $dW_X dW_Y = \rho dt$
Barlow (2002)	$S_t = \begin{cases} (1 + \alpha X_t)^{1/\alpha}, & 1 + \alpha X_t > \varepsilon_0 \\ \varepsilon_0^{1/\alpha}, & 1 + \alpha X_t \leq \varepsilon_0 \end{cases}$ $dX_t = -\lambda(X_t - a)dt + \sigma dW_t$
Villaplana (2003)	$\ln S_t = h(t) + X_t + Y_t$ $dX_t = -\kappa_X X_t dt + \sigma_X dW_1 + J_u dN(\lambda_u) - J_d dN(\lambda_d)$ $dY_t = -\kappa_Y (\mu - Y_t)dt + \sigma_Y dW_2$ $dW_1 dW_2 = \rho dt$
Geman / Roncoroni (2006)	$S(t) = \exp\{E(t)\}$ $dE(t) = [h(t) + \theta(\mu(t) - E(t^-))]dt + \sigma dW(t) + f(E(t^-))dJ(t)$
Hikspoors / Jaimungal (2007)	$S_t = \exp\{h(t) + X_t + J_t\}$ $dX_t = \lambda_X (Y_t - X_t)dt + \sigma_X dW_t$ $dY_t = \lambda_Y (\phi - Y_t)dt + \sigma_Y dZ_t$ $dJ_t = -\kappa J_t dt + dQ_t$
Benth / et al. (2007)	$S(t) = h(t) + X(t)$ $X(t) = \sum_{i=1}^n Y_i(t)$ $dY_i(t) = -\lambda_i^{-1} Y_i(t)dt + dL_i(t), \quad Y_i(0) = y_i$

Table 1.1: Models for electricity prices.

Using maximum likelihood estimation based on Kalman filtering he reports empirical results on electricity data from the California market.

Given observations which derive from hidden state process u_1, u_2, \dots, u_t , one may distinguish ‘online’ methods from ‘batch’ methods. The online methods provide an estimate \hat{u}_t which can easily be updated given an additional observation v_{t+1} ; while the batch methods estimate the whole series

(u_1, \dots, u_t) from (v_1, \dots, v_t) .

Following most of the literature in this thesis, we have used online methods (such as the Kalman filter and particle filter), rather than batch methods such as the EM algorithm. One motivation for doing this is the need of market participants to update their models in real time.

As is clear from the survey above, many authors have attempted to design models which capture the typical properties of electricity prices, namely seasonalities, spikes, and stochastic mean-reversion, but none of the models proposed so far has commanded wide assent.

In this thesis we consider some relatively simple multi-factor models, with a relatively small parameters. A second emphasis is the use of both spot and forward prices for parameter estimation, and a third is the use of filtering techniques to estimate the hidden processes, and hence the model parameters.

We will introduce three different spot price models from which we can also extract the futures prices. All of these models are capable of capturing some of the features of the spot price dynamics and imply certain dynamics for futures prices.

The first model (MROU model) is a Gaussian two-factor model where the spot prices is an mean reversion process which reverts to a stochastic mean, also fluctuating as Ornstein-Uhlenbeck process.

Since the presence of spikes is a fundamental feature of electricity prices, and any relevant spot price model should take this feature into account, we introduce a second model, an extension of the MROU model with a jump component (MROU with jumps).

However, the inclusion of the jump component introduces two kinds of problems in parameter estimation. The first is that the inclusion of jumps adds several new parameters, to describe the jump frequency and distribution. The second is that the jump models are non-Gaussian, and the best known filtering technique for these models, the particle filter, is not easily adapted to handle parameter estimation.

To avoid this problem, we introduce the third model (NLMROU model) based on the MROU model that produce spikes introducing only one more parameter.

In general terms, a statistical model is good if it successfully captures the main features of the observed data. Various statistical tests (such as the Kolmogorov-Smirnov goodness of fit) can be used to test and compare statistical models.

From this viewpoint, the theory of electricity prices is rather undeveloped. No systematic comparison of the various models in the literature has been made. One reason for this is that in many cases authors have proposed models, but have not yet developed techniques for parameter estimation.

In this thesis, we have followed other workers in this area in using fairly simple tests for our models. We have compared moments and sample paths of simulated and real data. Even these simple tests indicate that our models do not capture all the features of real prices.

Chapter 2

Filtering

2.1 State space formulation

A state space model is a representation of the joint dynamic evolution of an observable random vector v_t and a unobservable state vector u_t . It is based on two important sets of system equations: the measurement equation and the transition equation. The transition equation describes the evolution of the state vector and the measurement equation reflects how the state interacts with the vector of observations. The evolution of the state is assumed to be autonomous, that is it does not depend on the measurement equation.

We consider the non-linear and non-Gaussian state-space model, which is represented in the following general form: For $t = 1, 2, \dots, n$,

$$u_t = f_t(u_{t-1}, q_{t-1}) \quad (\text{transition equation}), \quad (2.1)$$

$$v_t = h_t(u_t, r_t) \quad (\text{measurement equation}), \quad (2.2)$$

where $f_t : \mathbb{R}^{n_u} \times \mathbb{R}^{n_q} \mapsto \mathbb{R}^{n_u}$ and $h_t : \mathbb{R}^{n_u} \times \mathbb{R}^{n_r} \mapsto \mathbb{R}^{n_v}$ are vector functions, which are assumed to be known and possibly non-linear. The process and measurement noises, q_t and r_t respectively, are independent with known but arbitrary densities. In addition, we assume that the initial distribution of u_0 is available, that is $p(u_0) := p(u_0|v_0)$.

Associated with a state-space model is the problem of estimating the unobservable state using a set of observations. To do so, from a Bayesian perspective we need to estimate the filtering density $p(u_t|v_{1:t})$, where $v_{1:t} := \{v_1, v_2, \dots, v_t\}$ is the past history of the observed process up to time t . If possible, we wish to do this recursively, so that, $p(u_t|v_{1:t})$ can be calculated by updating the estimate $p(u_t|v_{1:t-1})$ with the new observation v_t . The esti-

mate of the filtering density can be obtained in two stages (prediction and updating) as follows.

Applying the Chapman-Kolmogorov equation we can write the time update iteration as:

$$\begin{aligned} p(u_t|v_{1:t-1}) &= \int p(u_t|u_{t-1}, v_{1:t-1})p(u_{t-1}|v_{1:t-1})du_{t-1} \\ &= \int p(u_t|u_{t-1})p(u_{t-1}|v_{1:t-1})du_{t-1} \end{aligned} \quad (2.3)$$

by using in the last equation the Markov property. Now, after the observation v_t is available we use the Bayes rule to have

$$\begin{aligned} p(u_t|v_{1:t}) &= \frac{p(v_{1:t}|u_t)p(u_t)}{p(v_{1:t})} \\ &= \frac{p(v_t, v_{1:t-1}|u_t)p(u_t)}{p(v_t, v_{1:t-1})} \\ &= \frac{p(v_t|v_{1:t-1}, u_t)p(v_{1:t-1}|u_t)p(u_t)}{p(v_t|v_{1:t-1})p(v_{1:t-1})} \\ &= \frac{p(v_t|v_{1:t-1}, u_t)p(u_t|v_{1:t-1})p(v_{1:t-1})p(u_t)}{p(v_t|v_{1:t-1})p(v_{1:t-1})p(u_t)} \\ &= \frac{p(v_t|u_t)p(u_t|v_{1:t-1})}{p(v_t|v_{1:t-1})}, \end{aligned} \quad (2.4)$$

where the denominator could be written as

$$p(v_t|v_{1:t-1}) = \int p(v_t|u_t)p(u_t|v_{1:t-1})du_t. \quad (2.5)$$

Unfortunately, in general there do not exist closed-form expressions for equations (2.3) and (2.4). The main exception to this is where (2.1) and (2.2) are linear and the noise processes q_t , r_t are Gaussian, and in this case the solution is given by the Kalman Filter [48].

2.2 The Kalman filter

Equations (2.1) and (2.2) reduce to the following special case where a linear Gaussian state-space model is considered. To include a more general case we have included two additive components C_t and A_t to the transition and measurement equation respectively. We have

$$u_t = \underset{(n_u \times 1)}{C_t} + \underset{(n_u \times n_u)}{D_t} u_{t-1} + \underset{(n_u \times 1)}{q_{t-1}}, \quad (2.6)$$

$$v_t = \underset{(n_v \times 1)}{A_t} + \underset{(n_v \times n_u)}{B_t} u_t + \underset{(n_v \times 1)}{r_t}, \quad (2.7)$$

where the process and measurement noises are normally distributed

$$\begin{pmatrix} q_t \\ r_t \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_q & 0 \\ 0 & \Sigma_r \end{pmatrix} \right). \quad (2.8)$$

The initial condition becomes

$$u_0 \sim N(\bar{u}_0, \Sigma_0) \quad (2.9)$$

and the matrices C_t , D_t , A_t , B_t , Σ_q , and Σ_r are assumed to be known. $N(\mu, \Sigma)$ denotes a Gaussian density with mean μ and covariance Σ , that is:

$$N(\mu, \Sigma) := |2\pi\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right\}. \quad (2.10)$$

Here $|\cdot|$ denotes the determinant.

For the model above (2.6)-(2.9), it follows that the transition density $p(u_{t+1}|u_t)$ and the measurement density $p(v_t|u_t)$ are normal. It can be shown that this implies that also the prediction and filtering densities are normal, see [77] for details. We have

$$p(u_t|v_{1:t-1}) = N(u_{t|t-1}, \Sigma_{t|t-1}), \quad (2.11)$$

$$p(u_t|v_t) = N(u_{t|t}, \Sigma_{t|t}), \quad (2.12)$$

$$p(v_t|v_{1:t-1}) = N(\hat{v}_{t|t-1}, F_{t|t-1}), \quad (2.13)$$

where the conditional means $u_{t|t-1}$, $u_{t|t}$, $\hat{v}_{t|t-1}$ and conditional covariances $\Sigma_{t|t-1}$, $\Sigma_{t|t}$ and $F_{t|t-1}$ are computed by the following pseudo code of the Kalman filter.

Algorithm 1 (Kalman filter)

- **Step 1, Initialization**

Set

$$u_{0|0} = \bar{u}_0 \quad , \quad \Sigma_{0|0} = \Sigma_0,$$

and set $t = 1$.

- **Step 2, Prediction**

Compute

$$u_{t|t-1} = C_t + D_t u_{t-1|t-1},$$

$$\Sigma_{t|t-1} = D_t \Sigma_{t-1|t-1} D_t' + \Sigma_q.$$

- **Step 3, Innovation**

Define

$$e_t = v_t - \hat{v}_{t|t-1} \tag{2.14}$$

with

$$\hat{v}_{t|t-1} = A_t + B_t u_{t|t-1} \tag{2.15}$$

and compute

$$F_{t|t-1} = B_t \Sigma_{t|t-1} B_t' + \Sigma_r. \tag{2.16}$$

- **Step 4, Updating**

Compute

$$K_t = \Sigma_{t|t-1} B_t' F_{t|t-1}^{-1}, \tag{2.17}$$

$$u_{t|t} = u_{t|t-1} + \Sigma_{t|t-1} B_t' F_{t|t-1}^{-1} e_t, \tag{2.18}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} B_t' F_{t|t-1}^{-1} B_t \Sigma_{t|t-1}. \tag{2.19}$$

- **Step 5, Looping**

if $t < n$, set $t = t + 1$ and go to **Step 2**; else stop.

In the linear Gaussian case, the Kalman filter has strong optimality properties. (2.11)-(2.13) give the maximum likelihood estimator of u_t given $v_{1:t}$, and this is also the minimum mean square error estimator. That is, in the mean square sense no other algorithm can perform better than the Kalman filter in the Gaussian environment; see [40] for details.

In cases where the measurement or the transition equation are nonlinear, sub-optimal solutions such as the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) are commonly used to solve the problem. The extended Kalman filter simply linearize all nonlinear transformations and substitutes a Jacobian matrix for the linear transformations in the Kalman filter equations [41]. Although it is easy to implement it has a number of limitations especially if the system nonlinearities are severe or the true distribution is multimodal or highly skewed - see [75]. The unscented Kalman filter gives a more accurate performance for nonlinear equations, but does rely on the noise being Gaussian.

2.3 The unscented Kalman filter

An alternative filter with performance superior to the extended Kalman filter is the unscented Kalman filter.

Unlike the extended Kalman filter it does not approximate the nonlinear function of the process and the observation, it uses the true nonlinear models to approximate the distribution of the state variable u_t by applying an *unscented transformation* to it. The unscented transformation uses the so-called sigma-points that capture the mean and covariance of the original distributions and, when propagated through the true nonlinear system, capture the posterior mean and covariance accurately to third order. For more details see [47, 75]. Unlike the particle filter considered in the next section, which requires a large number of points, the unscented transform only requires $2n_u + 1$ points to capture the mean and covariance of a probability distribution in \mathbb{R}^{n_u} .

Let us consider a simplified version of the UKF formulation, where we as-

sume that both the transition and measurement noises are additive Gaussian, that is,

$$u_t = f(u_{t-1}) + q_{t-1}, \quad (2.20)$$

$$v_t = h(u_t) + r_t, \quad (2.21)$$

where $q_{t-1} \sim N(0, \Sigma_q)$ and $r_t \sim N(0, \Sigma_r)$. Here $u_t \in \mathbb{R}^{u_n}$, $v_t \in \mathbb{R}^{v_n}$.

The algorithm can be described as follows:

Algorithm 2 (Unscented Kalman filter)

- **Step 1, Initialization**

Set

$$\bar{u}_0 = E(u_0) \quad \text{and} \quad \Sigma_0 = E[(u_0 - \bar{u}_0)(u_0 - \bar{u}_0)'].$$

Set $t = 1$.

- **Step 2, Unscented transformation**

Compute the sigma points and weights.

$$\begin{aligned} i = 0 \quad & \chi_{t-1}(0) = \bar{u}_{t-1}, \\ & w_0^{(m)} = \frac{\lambda}{n+\lambda}, \\ & w_0^{(c)} = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta), \\ i = 1, \dots, n \quad & \chi_{t-1}(i) = \bar{u}_{t-1} + \left(\sqrt{(n+\lambda)\Sigma_{t-1}} \right)_i, \\ & w_i^{(m)} = w_i^{(c)} = \frac{1}{2(n+\lambda)}, \\ i = n+1, \dots, 2n \quad & \chi_{t-1}(i) = \bar{u}_{t-1} - \left(\sqrt{(n+\lambda)\Sigma_{t-1}} \right)_{i-n}, \\ & w_i^{(m)} = w_i^{(c)} = \frac{1}{2(n+\lambda)}. \end{aligned}$$

Here the subscripts i and $i-n$ correspond to the i^{th} and $i-n^{th}$ columns of the square-root matrix (Cholesky factorization). $\lambda = \alpha^2(n + \kappa) - n$ is a scaling parameter. α determines the spread of the sigma points around \bar{u}_{t-1} and is usually set to a small positive value. κ is a secondary

scaling parameter which is usually set to 0, and β is used to incorporate prior knowledge of the distribution of u_t .

- **Step 3, Time update**

For $i = 0, 1, \dots, 2n$,

$$\chi_{t|t-1}(i) = f(\chi_{t-1}(i))$$

$$v_{t|t-1}(i) = h(\chi_{t|t-1}(i))$$

$$\bar{u}_{t|t-1} = \sum_{i=0}^{2n} w_i^{(m)} \chi_{t|t-1}(i)$$

$$\Sigma_{t|t-1} = \sum_{i=0}^{2n_a} w_i^{(c)} (\chi_{t|t-1}(i) - \bar{u}_{t|t-1})(\chi_{t|t-1}(i) - \bar{u}_{t|t-1})' + \Sigma_q$$

$$\bar{v}_{t|t-1} = \sum_{i=0}^{2n} w_i^{(m)} v_{t|t-1}(i) \quad (2.22)$$

$$\Sigma_{v_t v_t} = \sum_{i=0}^{2n} w_i^{(c)} (v_{t|t-1}(i) - \bar{v}_{t|t-1})(v_{t|t-1}(i) - \bar{v}_{t|t-1})' + \Sigma_r \quad (2.23)$$

$$\Sigma_{u_t v_t} = \sum_{i=0}^{2n} w_i^{(c)} (\chi_{t|t-1}(i) - \bar{u}_{t|t-1})(v_{t|t-1}(i) - \bar{v}_{t|t-1})'.$$

- **Step 4, Measurement update**

Calculate

$$K_t = \Sigma_{u_t v_t} \Sigma_{v_t v_t}^{-1}, \quad (2.24)$$

$$u_t = \bar{u}_{t|t-1} + K_t(v_t - \bar{v}_{t|t-1}), \quad (2.25)$$

$$\Sigma_t = \Sigma_{t|t-1} - K_t \Sigma_{v_t v_t} K_t'. \quad (2.26)$$

- **Step 5, Looping**

if $t < n$, set $t = t + 1$, update u_t and Σ_t and go to **Step 2**; else stop.

2.4 Particle filter

A different approach to filtering has recently become popular [26, 79]. In this approach, we use Monte Carlo simulations instead of Gaussian approximations for $p(u_t|v_t)$, as in the Kalman filter. This method allows for a complete representation of the filtering distribution, so that any statistical estimate can be easily calculated. This filter has the advantage that it allows one to deal with fundamentally non-Gaussian situations.

The idea is based on the importance sampling technique [70, 76]. Recall that $u_{0:t} = \{u_0, u_1, \dots, u_t\}$ is the (unknown) true state of the system, and $v_{1:t} = \{v_1, v_2, \dots, v_t\}$ are the observations. Suppose we wish to calculate the conditional expectation

$$E(f(u_{0:t})|v_{1:t}) = \int f(u_{0:t})p(u_{0:t}|v_{1:t})du_{0:t}. \quad (2.27)$$

In most non-Gaussian or non-linear situations, the true distribution $p(u_{0:t}|v_{1:t})$ will be impossible to calculate. Importance sampling works by instead sampling from a proposal distribution $q(u_{0:t}|v_{1:t})$, which can be easily sampled from. The support of $q(u_{0:t}|v_{1:t})$ is assumed to cover that of $p(u_{0:t}|v_{1:t})$.

We can write

$$\begin{aligned} E(f(u_{0:t})|v_{1:t}) &= \int f(u_{0:t}) \frac{p(u_{0:t}|v_{1:t})}{q(u_{0:t}|v_{1:t})} q(u_{0:t}|v_{1:t}) du_{0:t}, \\ &= \int f(u_{0:t}) \frac{p(v_{1:t}|u_{0:t})p(u_{0:t})}{p(v_{1:t})q(u_{0:t}|v_{1:t})} q(u_{0:t}|v_{1:t}) du_{0:t}, \\ &= \int f(u_{0:t}) \frac{w(u_{0:t})}{p(v_{1:t})} q(u_{0:t}|v_{1:t}) du_{0:t}, \end{aligned} \quad (2.28)$$

where

$$w(u_{0:t}) = \frac{p(v_{1:t}|u_{0:t})p(u_{0:t})}{q(u_{0:t}|v_{1:t})} \quad (2.29)$$

is defined as the filtering non-normalized weight at step t . Now

$$\begin{aligned}
E(f(u_{0:t})|v_{1:t}) &= \frac{1}{p(v_{1:t})} \int f(u_{0:t})w(u_{0:t})q(u_{0:t}|v_{1:t})du_{0:t} \\
&= \frac{\int f(u_{0:t})w(u_{0:t})q(u_{0:t}|v_{1:t})du_{0:t}}{\int p(v_{1:t}|u_{0:t})p(u_{0:t})\frac{q(u_{0:t}|v_{1:t})}{q(u_{0:t}|v_{1:t})}du_{0:t}} \\
&= \frac{\int f(u_{0:t})w(u_{0:t})q(u_{0:t}|v_{1:t})du_{0:t}}{\int w(u_{0:t})q(u_{0:t}|v_{1:t})du_{0:t}} \\
&= \frac{E_q(f(u_{0:t})w(u_{0:t}))}{E_q(w(u_{0:t}))} \\
&= E_q(f(u_{0:t})\tilde{w}(u_{0:t})),
\end{aligned}$$

where

$$\tilde{w}(u_{0:t}) = \frac{w(u_{0:t})}{E_q(w(u_{0:t}))} \quad (2.30)$$

is defined to be the filtering normalized weight at step t .

Now let $u_{0:t}^{(i)}$, $i = 1, 2, \dots, n_p$ be a Monte-Carlo sample from $q(u_{0:t}|v_{1:t})$. Then by the law of large numbers

$$E(f(u_{0:t})|v_{1:t}) \approx \sum_{i=1}^{n_p} f(u_{0:t}^{(i)})\tilde{w}_t^{(i)} \quad (2.31)$$

where now

$$\tilde{w}_t^{(i)} = \frac{w(u_{0:t}^{(i)})}{\sum_{j=1}^{n_p} w(u_{0:t}^{(j)})}. \quad (2.32)$$

Note that (2.31) is correctly normalized: if $f \equiv 1$ then both sides of (2.31) equal 1.

Thus, provided we can sample from the proposal distribution $q(u_{0:t}|v_{1:t})$, and calculate the weights $w(u_{0:t})$ given by (2.29), then we can estimate $E(f(u_{0:t})|v_{1:t})$. Note also, that if we can only calculate $w(\cdot)$ to a multiplicative constant c , which can depend on $v_{1:t}$, then this constant drops out when we calculate \tilde{w}_t in (2.32), and so the procedure will still work.

Suppose that the importance density is chosen to factorize, so that

$$q(u_{0:t}|v_{1:t}) = q(u_t|u_{0:t-1}v_{1:t})q(u_{0:t-1}|v_{1:t-1}). \quad (2.33)$$

Then

$$\begin{aligned} w(u_{0:t}^{(i)}) &= \frac{p(v_{1:t}|u_{0:t}^{(i)})p(u_{0:t}^{(i)})}{q(u_{0:t}^{(i)}|v_{1:t})} \\ &= \frac{p(v_t, v_{1:t-1}|u_{0:t-1}^{(i)}, u_t^{(i)})p(u_t^{(i)}|u_{0:t-1}^{(i)})p(u_{0:t-1}^{(i)})}{q(u_t^{(i)}|u_{0:t-1}^{(i)}, v_{1:t})q(u_{0:t-1}^{(i)}|v_{1:t-1})}. \end{aligned}$$

The measurement equation (2.2) implies that v_t depends on $u_{0:t}$ only through u_t and that v_1, v_2, \dots, v_t are conditionally independent given $u_{0:t}$. Therefore

$$p(v_t, v_{1:t-1}|u_{0:t-1}^{(i)}, u_t^{(i)}) = p(v_t|u_t^{(i)})p(v_{1:t-1}|u_{0:t-1}^{(i)}).$$

Further, (2.1) implies that u_t is Markov, so

$$p(u_t^{(i)}|u_{1:t-1}^{(i)}) = p(u_t^{(i)}|u_{t-1}^{(i)}).$$

Combining these equations,

$$\begin{aligned} w_t^{(i)} &= \frac{p(v_{1:t-1}|u_{0:t-1}^{(i)})p(u_{0:t-1}^{(i)})p(v_t|u_t^{(i)})p(u_t^{(i)}|u_{t-1}^{(i)})}{q(u_{0:t-1}^{(i)}|v_{1:t-1})q(u_t^{(i)}|u_{0:t-1}^{(i)}, v_{1:t})} \\ &= w_{t-1}^{(i)} \frac{p(v_t|u_t^{(i)})p(u_t^{(i)}|u_{t-1}^{(i)})}{q(u_t^{(i)}|u_{0:t-1}^{(i)}, v_{1:t})}. \end{aligned} \quad (2.34)$$

Thus, the importance weights defined in (2.34) can be updated in a simple way at each time step. (2.31) implies that the filtering density $p(u_t|v_{1:t})$ can be approximated by

$$p(u_t|v_{1:t}) \approx \sum_{i=1}^{n_p} \tilde{w}_t^{(i)} \delta_{u_t^{(i)}}(u_t) \quad (2.35)$$

where $\delta_u(\cdot)$ denote the point mass at u . The success of this operation depends on how close the proposal distribution is to the posterior and whether the resulting point-mass approximation is an adequate representation of the distribution of interest. Although sequential importance sampling poses only one restriction on the importance density, equation (2.33), with the number of choices being unlimited, the design of the appropriate proposal function is, in fact, one of the most critical issues in importance sampling algorithms [79]. Poor choice leads to poor approximation in (2.35), and to poor algorithm performance in general. See [26, 69] for more details and variants of the particle filter.

One major problem with this algorithm is that the variance of the weights increases steadily over time. If one starts with a fixed number n_p of particles, then in practice after a while nearly all the mass of the distribution in (2.35) is concentrated at one particle. Not surprisingly, this leads to poor algorithm performance. In order to solve this, we resample the points to create copies of particles with large importance weights and to remove those with negligible importance weights. This ensures that there are sufficient particles exploring regions of high probability in the next time step [37]. Various methods have been suggested for this [2, 13, 58]. In the particle filter literature four basic resampling algorithms can be identified:

1. Multinomial resampling

Generate n_p ordered uniform random numbers

$$a_j = a_{j+1} \tilde{a}_j^{1/j}, \quad a_{n_p} = \tilde{a}_{n_p}^{1/n_p}, \quad \text{with } \tilde{a}_j \sim U[0, 1]$$

and use them to select the new particles $u^{(j)}$ according to the multino-

mial distribution. That is,

$$\begin{aligned} u^{(j)} &= u(F^{-1}(a_j)) \\ &= u^{(i)} \text{ with } i \text{ s.t. } a_j \in \left[\sum_{s=1}^{i-1} \tilde{w}_t^{(s)}, \sum_{s=1}^i \tilde{w}_t^{(s)} \right) \end{aligned}$$

where F^{-1} denotes the generalized inverse of the cumulative probability distribution of the normalized particle weights.

2. Stratified resampling

Generate n_p ordered random numbers

$$a_j = \frac{(j-1) + \tilde{a}_j}{n_p}, \text{ with } \tilde{a}_j \sim U[0, 1)$$

and use them to select $u^{(j)}$ according to the multinomial distribution.

3. Systematic resampling

Generate n_p ordered numbers

$$a_j = \frac{(j-1) + \tilde{a}}{n_p}, \text{ with } \tilde{a} \sim U[0, 1)$$

and use them to select $u^{(j)}$ according to the multinomial distribution.

4. Residual resampling

Allocate $n'_i = \lfloor n_p \tilde{w}_t^{(i)} \rfloor$ copies of particle $u^{(i)}$ to the new distribution. Additionally, resample $m = n_p - \sum n'_i$ particles from $\{u^{(i)}\}$ by making n''_i copies of particle $u^{(i)}$ where the probability for selecting $u^{(i)}$ is proportional to $\tilde{w}_t^{(i)} = n_p \tilde{w}_t^{(i)} - n'_i$ using one of the resampling schemes mentioned earlier.

A illustration of generic particle filter is shown in Figure 2.1. The whole particle filter algorithm can be implemented in the following way:

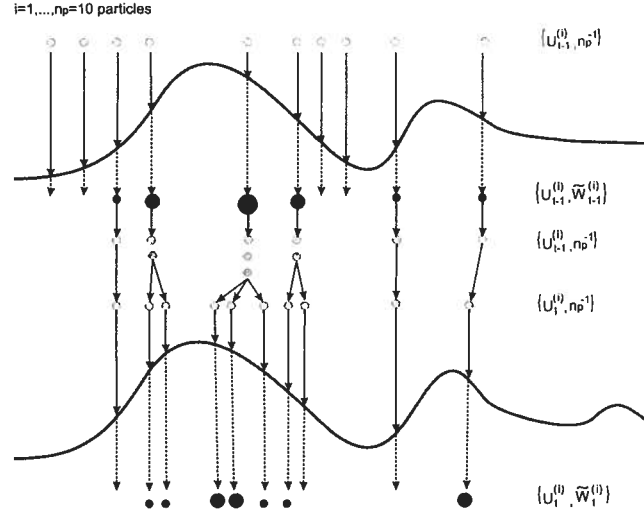


Figure 2.1: A graphical representation of the particle filter with importance sampling and resampling.

Algorithm 3 (Generic Particle Filter)

- **Step 1, Initialization**

For time step $t = 0$, choose u_0 and for each i between 1 and n_p (number of particles), take

$$u_0^{(i)} \sim p(u_0)$$

where $p(\cdot)$ is the initial distribution. Also take

$$w_0^{(i)} = \frac{1}{n_p} \quad (\text{importance weights}).$$

While $1 \leq t \leq n$ (number of observations)

- **Step 2, Prediction**

For each index i sample

$$u_t^{(i)} \sim q(u_t | u_{t-1}^{(i)}, v_{1:t}).$$

- **Step 3, Importance sampling**

Calculate the probabilities

$$\begin{aligned} p(v_t|u_t^{(i)}) & \quad (\text{likelihood distribution}), \\ p(u_t^{(i)}|u_{t-1}^{(i)}) & \quad (\text{prior distribution}), \\ q(u_t^{(i)}|u_{t-1}^{(i)}, v_{1:t}) & \quad (\text{proposal distribution}), \end{aligned} \tag{2.36}$$

and the associated weights for each i

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p(v_t|u_t^{(i)})p(u_t^{(i)}|u_{t-1}^{(i)})}{q(u_t^{(i)}|u_{t-1}^{(i)}, v_{1:t})}.$$

- **Step 4, Normalizing**

Normalize the weights

$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^{n_p} w_t^{(i)}}.$$

- **Step 5, Resampling**

Resample the points $u_t^{(i)}$ and reset $w_t^{(i)} = \tilde{w}_t^{(i)} = 1/n_p$.

- **Step 6, Looping**

Increment t and go back to **Step 2**. Stop at the end of the While loop.

2.5 Parameter estimation via maximum likelihood

Up to this point we have assumed that the functions f and h , with the distribution of the noise q_t, r_t are fully known. We have discussed the filtering problem - that is how to estimate $u_{0:t}$ from observations $v_{1:t}$. But in many applications, and in particular nearly all financial applications f, h, q_t , and r_t , will depend on unknown parameters. In this case the structure of the nonlinear state-space model becomes

$$u_t = f_t(u_{t-1}, q_{t-1}, \theta) \quad (\text{transition equation}), \quad (2.37)$$

$$v_t = h_t(u_t, r_t, \theta) \quad (\text{measurement equation}), \quad (2.38)$$

where $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ denotes the parameters in the model.

Given the above structure, in this section we address the problem of estimating the parameters θ from the observed data. The parameter estimation problem for state-space models has generated a lot of interest over the past few years and many techniques have been proposed to solve it. These methods could be broadly classified as Maximum Likelihood or Bayesian.

Using a Maximum Likelihood formulation the estimate of θ is the maximizing argument of the likelihood of the observed data, i.e.

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_\theta(v_1, v_2, \dots, v_t) \quad (2.39)$$

where $p_\theta(v_1, v_2, \dots, v_t)$ denotes the joint density of the observations up to time t . In a more convenient form we can rewrite (2.39) as

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}_\theta(v_{1:t}), \quad \mathcal{L}_\theta(v_{1:t}) = \log p_\theta(v_{1:t}) \quad (2.40)$$

where $v_{1:t} = \{v_1, v_2, \dots, v_t\}$. The joint density can be written as the product of the conditional densities:

$$p_\theta(v_{1:t}) = \prod_{k=1}^t p_\theta(v_k | v_{1:k-1}), \quad (2.41)$$

where $p_\theta(v_1 | v_0) = p_\theta(v_1)$. Thus the log-likelihood function is

$$\mathcal{L}_\theta(v_{1:t}) = \sum_{k=1}^t \log p_\theta(v_k | v_{1:k-1}). \quad (2.42)$$

The material presented up to here has dealt with the state-space model using a quite general formulation. The algorithm described above is in principle the full answer to the problem of parameter estimation. Now, from

a classical approach we can use some numerical optimization search procedure (Newton's method, Nelder-Mead) on (2.42) in order to calculate the maximum likelihood estimate $\hat{\theta}$.

Maximum likelihood estimation (MLE) of θ is particularly simple in the linear Gaussian state-space model (2.6)-(2.7), since the density function $p_{\theta}(v_k|v_{1:k-1})$ is the normal distribution with mean $\hat{v}_{k|k-1}$ and covariance matrix $F_{k|k-1}$ given by equations (2.15) and (2.16) respectively. Thus,

$$p_{\theta}(v_k|v_{1:k-1}) = |(2\pi)^{n_v}|F_{k|k-1}|^{-1/2} \exp\left\{-\frac{1}{2}(v_k - \hat{v}_{k|k-1})' F_{k|k-1}^{-1} (v_k - \hat{v}_{k|k-1})\right\}$$

and the log-likelihood function becomes

$$\mathcal{L}_{\theta}(v_{1:t}) \propto - \sum_{k=1}^t [\log |F_{k|k-1}| + (v_k - \hat{v}_{k|k-1})' F_{k|k-1}^{-1} (v_k - \hat{v}_{k|k-1})], \quad (2.43)$$

where $|F_{k|k-1}|$ denotes the determinant of $F_{k|k-1}$. Thus finding the MLE is quite straightforward for the Kalman filter.

In the general nonlinear case this approach is non-trivial, since the distribution $p_{\theta}(v_1, v_2, \dots, v_t)$ is generally not available in closed form. However, if we are using the UKF algorithm then, one can approximate the true (non Gaussian) distribution $p_{\theta}(v_k|v_{1:k-1})$ by a Gaussian distribution $N(\bar{v}_{k|k-1}, \Sigma_{v_k v_k})$, where $\bar{v}_{k|k-1}$ and $\Sigma_{v_k v_k}$ are given by equations (2.22) and (2.23). Hence one obtains

$$\mathcal{L}_{\theta}(v_{1:t}) \propto - \sum_{k=1}^t [\log |\Sigma_{v_k v_k}| + (v_k - \bar{v}_{k|k-1})' \Sigma_{v_k v_k}^{-1} (v_k - \bar{v}_{k|k-1})].$$

Since the UKF algorithm is based on the equations of a Kalman filter, the maximization of the log-likelihood can be done exactly as in the Kalman filter.

For the general nonlinear case, with non Gaussian noise, we have seen that the particle filter provides a technique for filtering with known parameters. In

general, we can approximate (2.42) using equation (2.5). Given the likelihood at step k ,

$$p_\theta(v_k|v_{1:k-1}) = \int p_\theta(v_k|u_k)p_\theta(u_k|v_{1:k-1})du_k.$$

this could be written as

$$p_\theta(v_k|v_{1:k-1}) = \int p_\theta(v_k|u_k) \frac{p_\theta(u_k|v_{1:k-1})}{q_\theta(u_k|u_{k-1}, v_{1:k})} q_\theta(u_k|u_{k-1}, v_{1:k}) du_k,$$

and given that by construction the $u_k^{(j)}$'s are distributed according to $q()$, we can write the Monte Carlo approximation

$$p_\theta(v_k|v_{1:k-1}) \approx \frac{1}{n_p} \sum_{j=1}^{n_p} w_k^{(j)}. \quad (2.44)$$

Thus, we can estimate $\mathcal{L}_t(\theta)$ by

$$\mathcal{L}_t(\theta) \approx \sum_{k=1}^t \log \left(\frac{1}{n_p} \sum_{j=1}^{n_p} w_k^{(j)} \right). \quad (2.45)$$

While this does give an approximation to the log-likelihood, this approach has several problems if we try to use it to obtain the MLE. For a fixed θ equation (2.45) gives a random variable, where the randomness comes from the particle filter. See Chapter 4 below for an account of the difficulties this causes.

An alternative approach to maximize the likelihood is employed the Expectation Maximization (EM) algorithm [74]. The objective of the algorithm is to maximize the likelihood of the observed data (2.42) in the presence of the hidden variables ($u_{0:t} = \{u_0, u_1, \dots, u_t\}$). The basic idea is that if we could observe $u_{0:t}$, in addition to the observations $v_{1:t}$ then we would consider $\{u_{0:t}, v_{1:t}\}$ as the complete data with the joint density

$$p_\theta(u_{0:t}, v_{1:t}) = p_\theta(u_0) \prod_{k=1}^t p_\theta(u_k|u_{k-1}) \prod_{k=1}^t p_\theta(v_k|u_k) \quad (2.46)$$

and seek the maximum log-likelihood estimate of θ via

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}_{\theta}(v_{1:t}, u_{0:t}), \quad \mathcal{L}_{\theta}(v_{1:t}, u_{0:t}) = \log p_{\theta}(v_{1:t}, u_{0:t}). \quad (2.47)$$

The EM algorithm for maximizing $\mathcal{L}_{\theta}(v_{1:t}, u_{1:t})$ is a two step procedure.

1. **(E-Step)** Computes the expected value of $\mathcal{L}_{\theta}(v_{1:t}, u_{1:t})$ over the hidden (missing) data $u_{1:t}$ based on the current value of the parameters $\theta^{(j)}$ and the observations $v_{1:t}$

$$\begin{aligned} Q(\theta|\theta^{(j)}) &= \mathbb{E}(\mathcal{L}_{\theta}(v_{1:t}, u_{0:t})|v_{1:t}, \theta^{(j)}) \\ &= \int \log p_{\theta}(u_{0:t}, v_{1:t}) p_{\theta^{(j)}}(u_{0:t}|v_{1:t}) du_{0:t}. \end{aligned}$$

2. **(M-Step)** Update the parameter estimate $\theta^{(j+1)}$ maximizing $Q(\theta|\theta^{(j)})$ with respect to θ ,

$$\theta^{(j+1)} = \arg \max_{\theta} Q(\theta|\theta^{(j)}) \quad (2.48)$$

and we repeat this two-step process until a fixed point of Q is obtained.

Unfortunately, there are very few situations where an exact and tractable solution exists for these two steps. One exception is the linear Gaussian state-space. The algorithm is described in [71]. Work applying the EM algorithm to nonlinear dynamical systems in the form of (2.20) and (2.21) is reported in [23, 34, 41].

To use the EM algorithm for a general nonlinear state-space the sequential Monte Carlo (particle filter) methods are employed to approximate $Q(\theta|\theta^{(j)})$ numerically. For the maximization step there is no standard method to solve this problem, and so it is necessary to proceed on case-by-case basis. In one general approach we calculate gradients with respect to θ , and use a gradient-based search procedure to find the maximum. Recent works that use this technique are in [36, 65, 81]. None of these methods is simple.

2.6 Parameter estimation via Bayesian methods

Because of these difficulties one would like an alternative to the MLE approximation. One Bayesian approach, described in ([43, 53, 57]) is to consider a Bayesian estimation by concatenating the state vector u_t with the unknown parameter θ , and introduce an artificial dynamic on the parameter. That is, we replace θ by θ_t and define a new state vector

$$y_t = \begin{pmatrix} u_t \\ \theta_t \end{pmatrix} \quad (2.49)$$

where $\theta_t \sim p(\theta_{t-1}|v_{1:t})$. Then one applies the particle filter to this augmented state-space. However this method has a number of difficulties and problems [50]. At a theoretical level it is not altogether satisfactory to replace a fixed parameter θ by a random evolution process θ_t . In addition, there are various practical problems associated with the choice of the artificial dynamic of θ . Thirdly, there are problems with the performance of the algorithm. These difficulties have been discussed in the literature, for example, in [18, 50].

Under the Bayesian framework, more sophisticated and new methods are being proposal. See [50] for an overview of particle filters methods for parameter estimation considering a nonlinear non-Gaussian state-space models.

Chapter 3

MROU model

In this chapter we present a Gaussian two-factor model known as Mean-Reverting to Ornstein-Uhlenbeck model (MROU) for the spot price and the convenience yield that captures some of the characteristics that we had described above of the power market and the dynamics of the futures prices.

3.1 Double mean-reversion model

There are many parallels between interest-rate models and modeling commodity prices, so many models originally developed for stock and interest rate markets have been applied to the energy market. We implement, for the valuation of electricity futures contracts, a two-factor mean-reverting model originally proposed in [5], and considered previously in [44] for oil commodity prices.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{Q})$ be a complete filtered probability space where \mathbb{Q} is the risk-neutral measure. If \tilde{S}_t is the spot price then

$$\tilde{S}_t = \exp\{X_t + h(t)\}, \quad (3.1)$$

$$dX_t = -\lambda_X(X_t - L_t) dt + \sigma_X dW_t^1, \quad (3.2)$$

$$dL_t = -\lambda_L(L_t - L) dt + \sigma_L dW_t^2. \quad (3.3)$$

Here X_t is the observed deseasonal log spot price, and L_t is a non-observed long-term mean process. We assume that both processes are given under the risk-neutral measure and the two Brownian motions, W_t^1 and W_t^2 satisfy $d\langle W_t^1, W_t^2 \rangle = \rho dt$. $h(t)$ is an arbitrary deterministic function that accounts for seasonality.

The difference between this model and the Gibson and Schwartz model

is that here, both the log spot price and the convenience yield follow an Ornstein-Uhlenbeck type process.

The seasonal component $h(t)$ combines trend and seasonality. Usually it consists of sum of sinusoidal functions which incorporate predictable daily and annual periodicity and dummy variables which incorporate predictable workday/weekend and holiday effects. In this thesis, we consider a sum of two cosine functions with distinct periods with a linear trend, that is

$$h(t) = \eta + \beta_0 t + \sum_{i=1}^2 \beta_i \cos\left(\frac{\tau_i + 2\pi t}{m_i}\right), \quad (3.4)$$

where m_i represent the seasonality period and the parameter

$$\phi = \{\eta, \beta_0, \beta_1, \beta_2, \tau_1, \tau_2\}$$

needs to be estimated.

The seasonalities have been discussed extensively in the literature, see for example Lucia and Schwartz [59], Cartea and Figueroa [16], and Benth et al. [7]. Although there are several ways of deseasonalising the data, usually it is estimated by means of non-linear regression methods. That is, if $t = t_1, t_2, \dots, t_n$ we estimate the seasonality function by fitting $h(t)$ to the log-prices using least squares estimation

$$\hat{\phi} = \arg \min_{(\eta, \beta_0, \beta_1, \beta_2, \tau_1, \tau_2)} \sum_{i=1}^n \left(h(t_i) - \log(\tilde{S}_{t_i}) \right)^2.$$

Write $\hat{h}(t)$ for the estimate. The deseasonalized log-spot price is given by:

$$X_t = \log \tilde{S}_t - \hat{h}(t).$$

3.2 Radon-Nikodym theorem for Ornstein-Uhlenbeck processes

Before we continue with parameter estimation for the MROU model, we consider a simple case, that is the problem of parameter estimation for the Ornstein-Uhlenbeck process $\text{OU}(\lambda, a, \sigma)$ defined by

$$dX_t = -\lambda(X_t - a)dt + \sigma dW_t. \quad (3.5)$$

If one observes the whole process $x = \{X_t, 0 \leq t \leq T\}$ the parameter σ can be estimated exactly, using the quadratic variation of X , but λ and a cannot.

The simplest statistical problem is when one has two alternatives:

$$H_0 : a = a_0, \lambda = \lambda_0 \quad \text{vs.} \quad H_1 : a = a_1, \lambda = \lambda_1$$

Denoting $\mathbb{P}_0, \mathbb{P}_1$ for the probability measures for (3.5) associated with H_0 and H_1 , the Radon-Nikodym theorem gives the optimal test for this in terms of the likelihood ratio. If

$$Z_T = \frac{L(x|H_1)}{L(x|H_0)} = \frac{d\mathbb{P}_1}{d\mathbb{P}_0},$$

then the test takes the form

$$\text{Reject } H_0 \text{ if } Z_T \geq c(\alpha),$$

where $c(\alpha)$ is given by

$$\mathbb{P}_0(\text{reject } H_0) = \mathbb{P}_0(Z_T \geq c(\alpha)) = \alpha.$$

The power of the test is given by

$$p = \mathbb{P}_1(\text{accept } H_1) = \mathbb{P}_1(Z_T \geq c(\alpha)).$$

Using the result in [44] (Thm. 3.2) we obtain

$$\begin{aligned} \log(Z_T) &= M_T - \frac{1}{2} \langle M \rangle_T \\ &= \int_0^T (c_0 + c_1 X_s) dW_s - \frac{1}{2} \int_0^T (c_0 + c_1 X_s)^2 ds. \end{aligned}$$

The constants c_0 , and c_1 satisfy

$$c_0 = \frac{\lambda_0 a_0 - \lambda_1 a_1}{\sigma}, \quad c_1 = \frac{\lambda_0 - \lambda_1}{\sigma}. \quad (3.6)$$

We used Monte Carlo simulation to find the power of the test for these sample tests - see Table 3.1.

	H-J 1	N-S 1	N-S 2	H-J 2
λ_0	0.73	5.78	5.78	0.73
a_0	4.21	4.83	4.83	4.21
λ_1	0.15	5.78	5.78	0.66
a_1	3.27	5.43	5.43	4.60
σ	0.63	1.03	1.03	0.63
T	3	1	0.5	25
α	0.10	0.05	0.10	0.10
p	100%	95%	86%	90%

Table 3.1: The data are taken from Hikspoors & Jaimungal (columns 1 and 4), and Nomikos & Soldatos (columns 2 and 3).

Given λ_0 , λ_1 , a_0 , a_1 , σ , and T we calculated c_0 , and c_1 from (3.6). We performed $n = 2000$ simulations of a standard OU(λ, a, σ) under \mathbb{P}_0 and \mathbb{P}_1 to estimate the power of the test.

Using these simulations, we can estimate how much data is needed to obtain reliable parameter estimation for the OU process.

We began by looking at some parameter values for OU processes found in the literature.

In [44] Hikspoors & Jaimungal study the MROU model (1.9). Having used futures data to estimate the parameters for X_t and Y_t in the risk neutral measure, and also estimate the process Y_t itself, they then estimate parameters for Y_t in the real world measure.

The column H-J 1 of Table 3.1 gives the values of those parameters. Considering these values, we have that the distribution of H_0 and H_1 are significantly different taking 3 years of observations, i.e. we distinguish (λ_0, a_0) from (λ_1, a_1) almost perfectly.

Nomikos & Soldatos [64] consider a regime switching model for the Nord Pool market - see (1.12)-(1.14). The parameters a_i , $i = 0, 1$ differ according to whether the weather is 'wet' or 'dry'. We can ask how long a period of observations is necessary to distinguish reliably between $a = a_0$ and $a = a_1$, in this situation. Simulation results shows that if $T = 1$ then this is possible at the 95% confidence level but if $T = 0.5$ then the power of the test is only

86% at $\alpha = 0.1$.

If we imagine data to be split into fixed periods in which the regime is constant and the task is to distinguish which regime holds during the period, then one can do so reliably if the period is 1 year, but shrinking the period to 6 months will give rise to an error probability of about 10%.

In the fourth column we consider two sets of parameters which differ by about 10%: the first set being similar to the values of (λ_0, a_0) in H-J 1.

In this case we see that even 25 years of data is not enough to reliably distinguish between the two parameter sets. Figure 3.1 plots the distribution of $\log Z_T$ under the two hypotheses: a substantial overlap in the distribution is apparent.

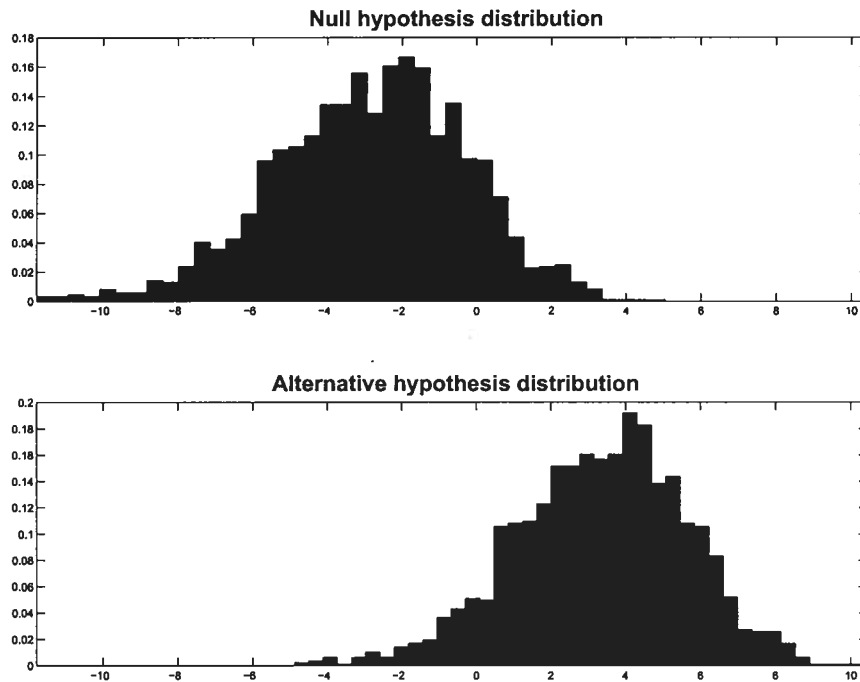


Figure 3.1: Sampling distributions with (column 4) where we add 10% deviation from Hikspoors & Jaimungal's parameters (column 1).

It would be interesting to perform an analysis using the Radon-Nikodym theorem for hypothesis testing with the MROU model. Calculating the likelihood Z_T with respect to the σ -field $\sigma(X_s, L_s, 0 \leq s \leq T)$ is straightforward - see [44] (Thm. 3.2). However, we need to calculate $\mathbb{E}(Z_T | X_s, 0 \leq s \leq T)$, and this would require calculating conditional expectations of the form

$$\mathbb{E} \left(\exp \left\{ \int_0^T (c_0 + c_1 L_s) dW_s - \frac{1}{2} \int_0^T (c_0 + c_1 L_s)^2 ds \right\} \middle| X_s, 0 \leq s \leq T \right).$$

This does not seem to be an easy problem.

However, from the results for the fourth column, it seems clear that many decades of data would be needed for an accurate estimate of the parameters for the hidden process L_t .

Financial time series have one significant property which makes them unique in a statistical sense. As well as the data itself (e.g. the spot prices), we also have available prices of various derivatives of the product. These produce a substantial amount of extra information.

In the context of the MROU model, we will therefore use futures as well as spot prices in our parameter estimation.

3.3 Future price

This model is an special case of Affine Jump-Diffusion model (AJD), so to obtain closed form formula for the price of futures contracts we use the results by Duffie et al. (2000) [27], see Appendix A.

The model can be rewritten as

$$\begin{aligned} d \begin{bmatrix} X_t \\ L_t \end{bmatrix} = & \begin{bmatrix} 0 \\ \lambda_L L \end{bmatrix} dt + \begin{bmatrix} -\lambda_X & \lambda_X \\ 0 & -\lambda_L \end{bmatrix} \begin{bmatrix} X_t \\ L_t \end{bmatrix} dt \\ & + \begin{bmatrix} \sqrt{1 - \rho^2} \sigma_X & \rho \sigma_X \\ 0 & \sigma_L \end{bmatrix} \begin{bmatrix} dW_t^1 \\ dW_t^2 \end{bmatrix}. \end{aligned} \quad (3.7)$$

Thus, $\mathbf{U}_t = [X_t, L_t]'$, and we have

$$d\mathbf{U}_t = (\mathbf{K}_0 + \mathbf{K}_1 \mathbf{U}_t)dt + d\mathbf{W}_t,$$

where $d\mathbf{W}_t$ is a Brownian motion with covariance \mathbf{H}_0 and

$$\mathbf{K}_0 = \begin{bmatrix} 0 \\ \lambda_L L \end{bmatrix}, \quad \mathbf{K}_1 = \begin{bmatrix} -\lambda_X & \lambda_X \\ 0 & -\lambda_L \end{bmatrix}, \quad \mathbf{H}_0 = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_L \\ \rho\sigma_X\sigma_L & \sigma_L^2 \end{bmatrix}. \quad (3.8)$$

The functions \mathbf{H}_1 and l_0 given in the Appendix A satisfy $\mathbf{H}_1 = \mathbf{0}$ and $l_0 = 0$. Duffie et al. gave expressions for various functionals of \mathbf{U} , and in particular for

$$\Psi(\mathbf{u}, t, T, \mathbf{U}_t) := \mathbb{E}[\exp\{\mathbf{u} \cdot \mathbf{U}_T\} | \mathbf{U}_t]. \quad (3.9)$$

By setting $\mathbf{u} = (1, 0)'$ in (3.9) one can obtain the future price of \mathbf{U}_t

$$\tilde{F}(t, T) := \Psi((1, 0)', t, T, \mathbf{U}_t) \exp\{h(T)\} \quad (3.10)$$

$$= \mathbb{E}[\exp\{X_T\} | \mathbf{U}_t] \exp\{h(T)\}. \quad (3.11)$$

By equation (A-3) in Appendix A, the future price is of the form

$$\begin{aligned} & \mathbb{E}[\exp\{X_T + h(T)\} | \mathbf{U}_t] \\ &= \exp\{h(T)\} \exp\{M((1, 0)', t, T) + N_1(1, t, T)X_t + N_2(0, t, T)L_t\}, \end{aligned} \quad (3.12)$$

where $M((1, 0)', t, T)$, $N_1(1, t, T)$ and $N_2(0, t, T)$ satisfy the following equations:

$$\frac{\partial N_1}{\partial t} = \lambda_X N_1, \quad (3.13)$$

$$\frac{\partial N_2}{\partial t} = -\lambda_X N_1 + \lambda_L N_2, \quad (3.14)$$

$$\frac{\partial M}{\partial t} = -\lambda_L L N_2 - \frac{1}{2}(\sigma_X^2 N_1^2 + \sigma_L^2 N_2^2) - N_1 N_2 \rho \sigma_X \sigma_L, \quad (3.15)$$

with the boundary conditions

$$M((1, 0)', T, T) = 0, \quad N_1(1, T, T) = 1, \quad \text{and} \quad N_2(0, T, T) = 0.$$

Solving the initial value problem, we have ⁵

$$N_1(t, T) = e^{\lambda_X(t-T)}, \quad (3.16)$$

$$N_2(t, T) = e^{\lambda_X(t-T)}m - e^{\lambda_L(t-T)}m, \quad (3.17)$$

$$\begin{aligned} M(t, T) = & (e^{\lambda_X(t-T)} - 1)m_1 + (e^{\lambda_L(t-T)} - 1)m_2 + (e^{2\lambda_X(t-T)} - 1)m_3 \\ & + (e^{(\lambda_X + \lambda_L)(t-T)} - 1)m_4 + (e^{2\lambda_L(t-T)} - 1)m_5, \end{aligned} \quad (3.18)$$

where

$$\begin{aligned} m = -\frac{\lambda_X}{\lambda_X - \lambda_L} \quad m_4 = \frac{m^2\sigma_L^2}{\lambda_X + \lambda_L} + \frac{m\rho\sigma_X\sigma_L}{\lambda_X + \lambda_L} \quad m_2 = Lm \\ m_1 = -\frac{\lambda_L Lm}{\lambda_X} \quad m_3 = -\left(\frac{\sigma_X^2}{4\lambda_X} + \frac{\sigma_L^2 m^2}{4\lambda_X} + \frac{m\rho\sigma_X\sigma_L}{2\lambda_X}\right) \quad m_5 = -\frac{m^2\sigma_L^2}{4\lambda_L}. \end{aligned}$$

Thus, we have that the price of the futures contract is given by

$$\tilde{F}(t, T) = \exp\{M(t, T) + N_1(t, T)X_t + N_2(t, T)L_t\} \exp\{h(T)\}, \quad (3.19)$$

where the functions $N_1(t, T)$, $N_2(t, T)$, and $M(t, T)$ are as (3.16)-(3.18). The deseasonalized log-future price is given by

$$\log F(t, T) = \tilde{F}(t, T) \exp\{-\hat{h}(T)\}. \quad (3.20)$$

3.4 Formulation in Kalman filter terms

We assume that data are available in the form of the spot price S_t and various futures or forward prices $F(t, T_i)$, $i = 1, 2, \dots, m$. Most data sets are available in the form of daily prices. Spot prices are traded for all hours of the week, but financial markets, which trade future contracts, are only open Monday-Friday. Since in any case, due to low consumption at the weekend (see Figure

⁵(see Appendix B for the calculations)

1.4b), spot prices on Saturdays and Sundays exhibit a different behavior to the rest of the week, we will disregard weekends and also holidays, and just consider weekday data. Given this data, we wish to estimate the parameters λ_X , λ_L , σ_X , σ_L , L and ρ .

The Kalman filter method has been applied previously to electricity models [14, 51]. To use the Kalman filter we need a discrete time set of equations, so we replace (4.2)-(4.4) with the forward Euler approximation:

$$S_t = e^{X_t}, \quad (3.21)$$

$$X_t = X_{t-1} - \lambda_X(X_{t-1} - L_{t-1})\Delta t + \sqrt{1 - \rho^2}\sigma_X\Delta W_t^1 + \rho\sigma_X\Delta W_t^2, \quad (3.22)$$

$$L_t = L_{t-1} - \lambda_L(L_{t-1} - L)\Delta t + \sigma_L\Delta W_t^2. \quad (3.23)$$

Here we have made a slight abuse of notation, in writing X_t, X_{t-1} for the successive values of X . More precisely, in (3.22) we should write

$$X_{t_j} = X_{t_{j-1}} - \lambda_X(X_{t_{j-1}} - L_{t_{j-1}})\Delta t + \sqrt{1 - \rho^2}\sigma_X\Delta W_{t_j}^1 + \rho\sigma_X\Delta W_{t_j}^2,$$

but to avoid two levels of subscripts we have used the form (3.22)-(3.23).

Here $\Delta t = 1/250$ (the number of trading days in a year), and $\Delta W_t^1, \Delta W_t^2$ are independent Gaussian random variables with mean 0 and variance Δt .

To apply the Kalman filter, the model must be expressed in its state-space form. Taking the state variable as $u_t = (X_t, L_t)'$ a discretization of the time $t = t_1, t_2, \dots, t_n$, and $\Delta t = (t_i - t_{i-1})$ the transition equation becomes:

$$u_t = C_t + D_t u_{t-1} + q_{t-1}, \quad (3.24)$$

where

$$C_t = \begin{pmatrix} 0 \\ \lambda_L L \Delta t \end{pmatrix}, \quad D_t = \begin{pmatrix} 1 - \lambda_X \Delta t & \lambda_X \Delta t \\ 0 & 1 - \lambda_L \Delta t \end{pmatrix},$$

and the process noise covariance matrix is

$$\Sigma_q = \text{cov}(q_{t-1}) = \begin{pmatrix} \sigma_X^2 \Delta t & \rho \sigma_X \sigma_L \Delta t \\ \rho \sigma_X \sigma_L \Delta t & \sigma_L^2 \Delta t \end{pmatrix}. \quad (3.25)$$

The measurement equation is given by the functions M , N_1 , and N_2 calculated according to equations (3.16)-(3.18):

$$v_t = \begin{pmatrix} X_t^* \\ Z_{t,T_i} \end{pmatrix} = A_t + B_t \begin{pmatrix} X_t \\ L_t \end{pmatrix} + r_t. \quad (3.26)$$

We write X_t^* for the observed deseasonalized log-spot price, and Z_{t,T_i} for the observed deseasonalized log price at time t of a future contract with maturity $T_i > t$. We will assume that there is some noise in the measurement of X_t^* and Z_{t,T_i} , so that

$$X_t^* = X_t + \xi_t^1, \quad (3.27)$$

$$Z_{t,T_i} = \log F(t, T_i) + \xi_{t,i}^2. \quad (3.28)$$

Here ξ_t^1 are iid $N(0, \sigma_S^2)$ random variables and $\xi_{t,i}^2$ are iid $N(0, \sigma_{Z_i}^2)$ random variables. We have two reasons for making this assumption about non-zero noise. First, at a fundamental level, it is reasonable to allow for some pricing errors due to large bid-ask spreads. (This may be particularly relevant in the futures markets, which are not always heavily traded.) Secondly, the Kalman filter involves matrix inversion - see (2.17), where $F_{t|t-1}^{-1}$ has to be computed. If the model has degeneracy, then severe numerical problems can arise. Adding the noise in (3.27) and (3.28) avoids this difficulty.

Using (3.19) we can write (3.28) as

$$Z_{t,T_i} = M(t, T_i) + X_t N_1(t, T_i) + L_t N_2(t, T_i) + \xi_{t,i}^2. \quad (3.29)$$

The measurement equation is therefore given by

$$v_t = \begin{pmatrix} X_t^* \\ Z_{t,T_1} \\ \vdots \\ Z_{t,T_m} \end{pmatrix} = A_t + B_t \begin{pmatrix} X_t \\ L_t \end{pmatrix} + r_t, \quad (3.30)$$

where

$$A_t = \begin{pmatrix} 0 \\ M(t, T_1) \\ \vdots \\ M(t, T_m) \end{pmatrix}, \quad B_t = \begin{pmatrix} 1 & 0 \\ N_1(t, T_1) & N_2(t, T_1) \\ \vdots & \vdots \\ N_1(t, T_m) & N_2(t, T_m) \end{pmatrix}, \quad (3.31)$$

and the measurement noise r_t has covariance matrix

$$\Sigma_r = \text{cov}(r_t) = \begin{pmatrix} \sigma_S^2 & 0 & \cdots & 0 \\ 0 & \sigma_{Z_1}^2 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{Z_m}^2 \end{pmatrix}. \quad (3.32)$$

Simulated trajectories of the MROU model using equations (3.24) and (3.26) with parameters $\lambda_X = 130$, $\lambda_L = 3$, $\sigma_X = 5$, $\sigma_L = 0.5$, $L = 3.5$, $\rho = 0.3$ and $\Delta t = 1/250$ can be seen in Figure 3.2. The long-term mean process L_t reverts towards the mean L , and as we expected, the spot price S_t and the future price $F(t, T_1)$ mimic the long-term mean process but with different volatility.

The observation and state equation matrices C_t , D_t , Σ_q , A_t , B_t , and Σ_r depend on the unknown parameters of the model. Based on this state-space formulation the parameters that we need to estimate are:

$$\theta = \{\lambda_X, \lambda_L, \sigma_X, \sigma_L, L, \rho, \sigma_S, \sigma_{Z_i}\}.$$

Note that if we use two different maturity contracts then $m = 2$, and we will have σ_{Z_1} and σ_{Z_2} in the parameter vector θ .

The log-likelihood function $\mathcal{L}_\theta(v_{1:t})$ for the linear Gaussian space-state is given by equation (2.43). This function can be maximized with respect to θ using an appropriate numerical optimization procedure. $\mathcal{L}_\theta(v_{1:t})$ only depends on the prediction error e_t and its covariance matrix $F_{t|t-1}$. Both in turn are outputs of the Kalman filter, equations (2.14) and (2.16). Thus the maximum likelihood estimate of θ can be obtained as follows:

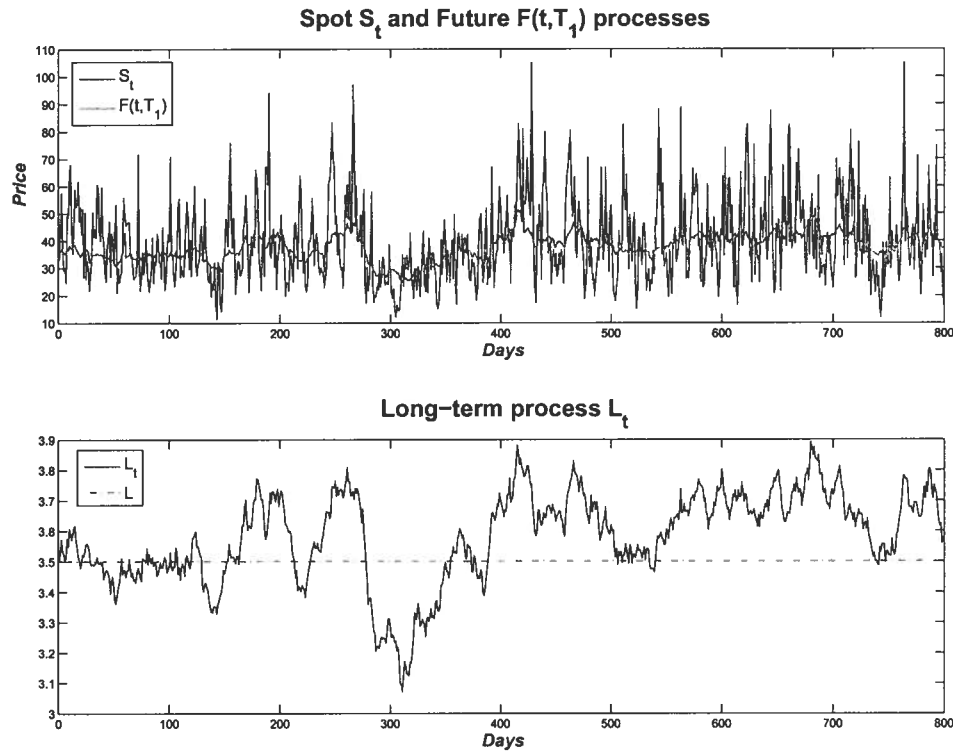


Figure 3.2: The upper graph shows the simulated spot price S_t and the future price $F(t, T_1)$ with maturity of one month. The lower graph is the long-term mean process L_t .

Algorithm 4 (Kalman Filter optimization)

- **Step 1**
Choose a initial value for θ , say θ_0 .
- **Step 2**
Run the Kalman Filter (Algorithm 1) and use the sequences e_t and $F_{t|t-1}$ to compute the log-likelihood $\mathcal{L}_\theta(v_{1:t})$ by (2.43).
- **Step 3**
Employ an optimization procedure that repeats **Steps 1-2** until a maximizer $\hat{\theta}$ of (2.43) has been found.

Some practical problems arise with the optimization procedure, and so the performance and accuracy of the Kalman Filter are affected since the problem may be poorly scaled. An optimization problem is poorly scaled if changes in the decision variables produce large changes in the objective function for some components and not for others [25]. We solved this problem by rescaling the variables.

There are several numerical search algorithms available to maximize the log-likelihood (**Step 3**). Barlow et al. [4] used the Nelder-Mead method⁶ to minimize $\mathcal{L}_\theta(v_{1:t})$. Here we decided to apply a quasi-Newton algorithm, the so-called BFGS method⁷ (which is similar to the method used by [51]) to get the initial point, and then used the Nelder-Mead method to find the optimum.

3.5 Empirical results

In this section we report some empirical results based on simulated and real data to examine the Kalman filter method applied to the model described in the previous section.

3.5.1 Simulated data

We first ran the algorithm on deseasonalized simulated data. To analyze the algorithm performance, first we simulated series with 100 and 800 observations respectively using equations (3.24) and (3.26), considering only the nearest monthly futures contracts in the log-future price, i.e. Z_{t,T_1} with $T_1 = 30$ days. In this case we have a 8 dimensional parameter space. We started the maximization procedure with a different initial values each time. Examples of some runs are given in Table 3.2.

Since we are searching for the maximum likelihood, we did 25 runs and we took the one which gave the largest value for $\mathcal{L}_t(\theta)$. Repeating the same

⁶The Nelder-Mead method is used to minimize a function of multiple variables without derivatives [see Nelder and Mead (1965)].

⁷Details and derivation of the BroydenFletcherGoldfarbShanno method in the context of filtering can be found in [28].

Chapter 3. MROU model

$n = 300$						
	True Value	Run 1	Run 2	Run 3	Run 4	Run 5
λ_X	130	129.032	127.923	126.335	138.586	130.232
λ_L	3	3.478	3.301	2.348	2.788	2.985
σ_X	5	4.917	4.845	4.658	4.795	5.123
σ_L	0.5	0.522	0.489	0.453	0.522	0.509
L	3.5	3.495	3.509	3.503	3.499	3.497
ρ	0.3	0.318	0.313	0.247	0.382	0.311
σ_S	0	0.004	0	0.005	0.000	0
σ_{Z_1}	0	0.009	0.002	0	0.000	0
$-\mathcal{L}_t(\theta)$		-620.952	-629.702	-643.223	-631.359	-601.72
CPU time		163.141	192.954	152.183	155.216	123.2167

Table 3.2: Five different maximization runs, on the same set of simulated data.

$n = 1000$			$n = 300$		
	True value	Estimator	Std.	Estimator	Std.
λ_X	130	129.369	1.055	129.042	4.221
λ_L	3	2.954	0.237	2.968	0.643
σ_X	5	4.998	0.098	4.862	0.127
σ_L	0.5	0.493	0.004	0.489	0.023
L	3.5	3.507	0.001	3.502	0.009
ρ	0.3	0.308	0.018	0.301	0.077
σ_S	0	0.0003	0.0007	0.001	0.002
σ_{Z_1}	0	0	0.003	0	0.001

Table 3.3: Estimation using one futures contract (average of 50 simulations).

procedure with 50 different series we obtained the following results, which are given in Table 3.3.

We can see that the estimation results recover the true values reasonably well in both cases. As expected, the standard deviation increases for all estimators when we reduce the number of days. Note that the estimator is close to zero for variables σ_S and σ_{Z_1} . This is not surprising since we are estimating the true model that generated the data and the noise in the model only comes from two sources.

Next, we simulated 25 new series with 300 data, but this time we included

log-futures prices with two different maturities, $T_1 = 30$ and $T_2 = 60$. In this case we have an extra parameter to estimate, σ_{Z_2} . In Table 3.4 we summarize the results, repeating the same procedure as before.

$n = 300$			
	True value	Estimator	Std.
λ_X	130	129.801	0.436
λ_L	3	3.000	0.000
σ_X	5	4.845	0.178
σ_L	0.5	0.499	0.001
L	3.5	3.505	0.017
ρ	0.3	0.294	0.102
σ_S		0.010	0.009
σ_{Z_1}		0.001	0
σ_{Z_2}		0	0

Table 3.4: Estimation using two futures contracts (n=300).

Thus we obtained good approximations for all the parameters using only a short data series. This could be useful due to the scarcity of data in the electricity markets. One example is [16] where the authors comment that there is too little data for parameter estimation in the UK market. We remark that if we tried to estimate the model parameters just using the spot price, then it will require many decades of data to make accurate estimates.

Notice that we used an alternative Kalman filter formulation (U-D filtering) since numerical problems arise when we include more parameters to estimated. During one of the recursions of the filter the covariance matrix failed to be positive semi-definite and consequently the estimated parameter differed from the true values. This problem arose because the matrices Σ_r and Σ_q were ill-conditioned. Since we are minimizing a fixed function, it is legitimate to discard runs which fail in this fashion. For further details about the U-D filter refer to [15, 38, 75].

3.5.2 The German electricity market

We now wish to apply these techniques to real data. While there are many markets in which electricity is traded, the data we require (that is, both spot and futures prices) in many cases are not available. For example, the Alberta

Power Pool makes spot prices available, but forward prices are known only to market participants. One market for which both sets of data are available is the German EEX market.

The European Energy Exchange (EEX) is Germany's energy exchange. It is one of the biggest power markets in Europe. EEX emerged in 2002 from the merger of EEX Leipzig Power Exchange and EEX European Energy Exchange Frankfurt. Both exchanges initially started spot trading for physical contracts in 2000. In 2001 EEX Frankfurt also introduced trading of standardized financial contracts. Commonly traded products in the power markets are baseload, peakload and hourly contracts. At the German market the times for peakload are defined as weekdays between 8:00am and 8:00pm. In the futures market contracts on both baseload and peakload are traded. The usual delivery periods are one month, one quarter and one year. In the spot market of the EEX baseload, peakload and hourly contracts up to the next weekday are traded.

The estimates are based on historical daily average spot price and monthly baseload futures price covering the period from July 2002, when the EEX and LPX markets merged, until the end of June 2007, almost five years of historical data. This data contains prices for 1267 days.

Figure 3.3 depicts the price trajectories of the spot and the nearby monthly futures prices between July 1, 2002 and June 29, 2007 for the EEX market. From the graph we note that there is a strong mean reversion and the spot prices show extreme spikes as well as high volatility which changes rapidly over short time periods. Moreover there is an linear drift over the years in the spot and futures prices.

Figure 3.3 shows much greater volatility, and more price spikes, for the period Jan 2005-June 2007 than in the earlier period, Jul 2002-Dec 2004. We therefore split the data into two parts (7/1/02-12/31/04, 1/1/05-6/29/07), and repeated the runs on each of these.

Following [8], we removed the seasonality by representing it as a linear combination of cosines including a trend, a weekly, and a yearly cycle of the form

$$h(t) = \eta + \beta_0 t + \beta_1 \cos\left(\frac{\tau_1 + 2\pi t}{250}\right) + \beta_2 \cos\left(\frac{\tau_2 + 2\pi t}{5}\right). \quad (3.33)$$

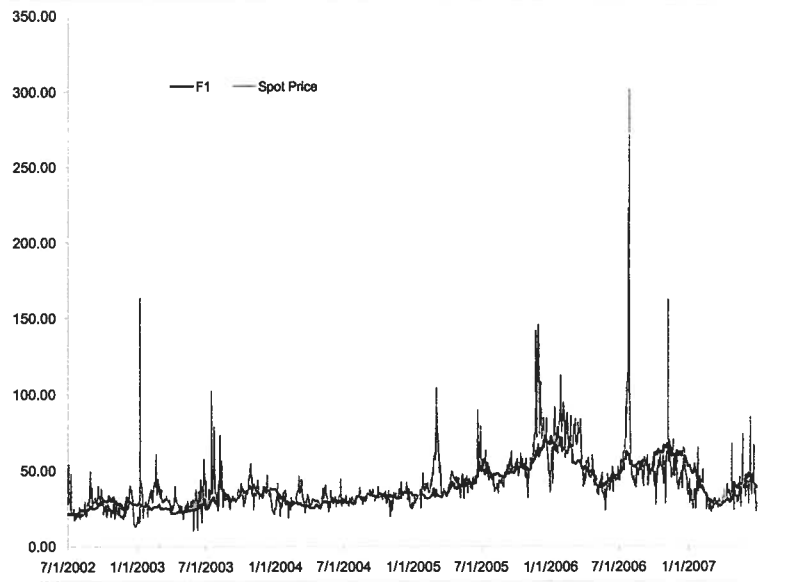


Figure 3.3: Electricity spot and nearby monthly futures price in German market.

Assuming 250 trading days in a year, we estimate the seasonality by fitting the $h(t)$ function to the log-price series by ordinary least squares. The results can be seen in Table 3.5 and Figures 3.4, and 3.5. We see less strong seasonality in the EEX market that in the more hydro-dependent Nordic electricity market that depends more on hydro. The seasonal component is highest in winter.

Parameter	$\hat{\eta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\tau}_1$	$\hat{\beta}_2$	$\hat{\tau}_2$
Est. value	3.2965	0.0005	-0.0695	23.5235	0.0224	0.8218

Table 3.5: Estimated values for $f(t)$ by least-squares fitting.

Based on equations (3.24) and (3.26) we estimated the model using the spot price and one futures contract with one month maturity, for both periods. Also we estimated using two futures contracts with one and two month

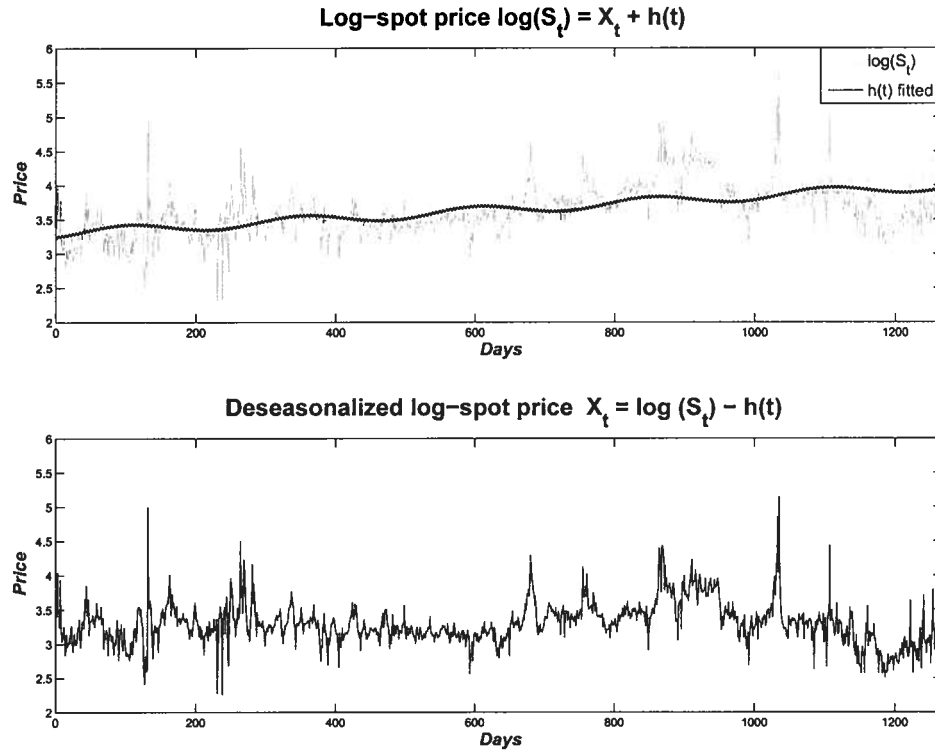


Figure 3.4: The upper graph shows the log spot-price of the EEX market and the seasonal component $\hat{h}(t)$ and the lower graph the deseasonal series $X_t = \log \tilde{S}_t - \hat{h}(t)$.

maturities. The results of the parameter estimation are shown in Tables 3.6 and 3.7.

For all the periods the estimates of ρ are quite small - the largest value being 0.115 for Part 1 when estimated using one future contract. This may be compared with the estimate -0.96 obtained in [44] in the context of the oil market.

Unlike the case of simulated data, the parameter estimates using two futures prices differ somewhat from those obtained with just one future. The explanation is presumably that the model does not perfectly describe the real data.

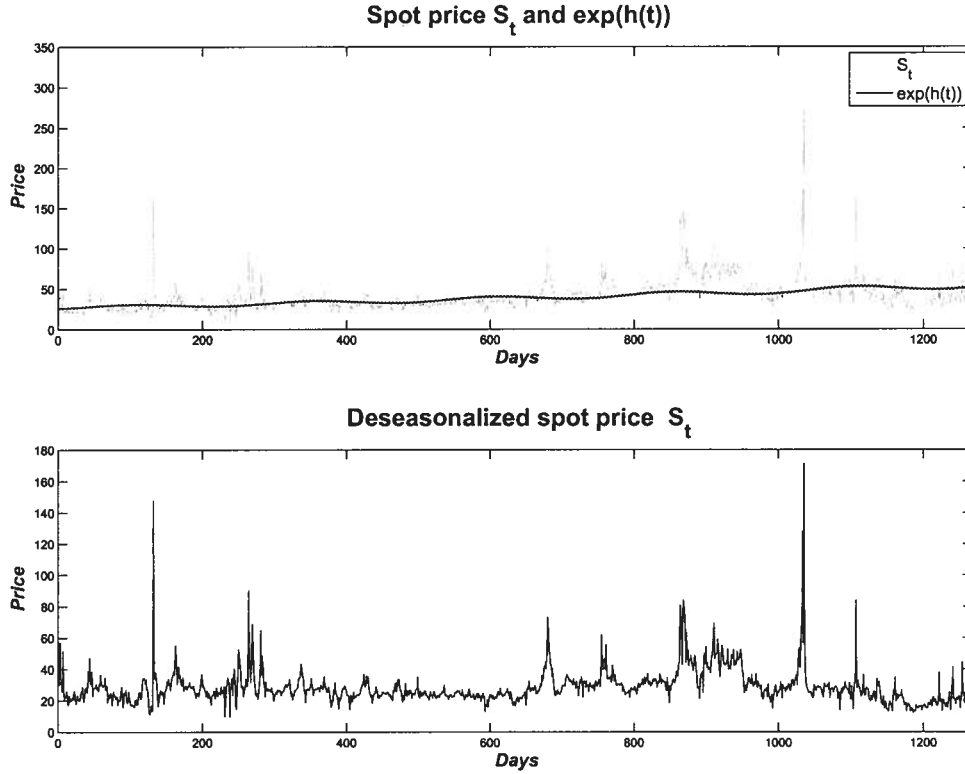


Figure 3.5: The upper shows the spot price \tilde{S}_t with $\exp\{\hat{h}(t)\}$ and the lower graph the deseasonalized spot price $S_t = \exp\{X_t\}$.

In the absence of noise, the model (4.2)-(4.4) gives exponential decay toward the long run mean L . We can therefore interpret

$$\tau_X = \frac{\log 2}{\lambda_X}, \quad \tau_L = \frac{\log 2}{\lambda_L}$$

as the ‘half lives’ of the processes X_t and L_t (measured in years). For the X_t process the estimates in Tables 3.6 and 3.7 give a half life 1.5 – 4 days, while for the L_t process the half life estimates vary from about 6–12 months. Thus the estimated process X_t and L_t do play a satisfactory role in separating out short and long-term fluctuation in the spot price.

The long run standard deviation of the OrnsteinUhlenbeck process L_t is $(\sigma_L/2\lambda_L)^{1/2}$. Using the numerical values for the whole period from Table 3.7,

	Whole	Part 1	Part 2
	Estimator	Estimator	Estimator
λ_X	65.030	42.439	77.509
λ_L	0.544	1.319	1.060
σ_X	1.719	1.172	2.013
σ_L	0.517	0.410	0.586
L	3.108	3.105	3.156
ρ	0.101	0.115	0.065
σ_S	0.177	0.193	0.167
σ_{Z_1}	0.000	0.000	0.000
CPU time	613.88	485.99	312.33

Table 3.6: Estimated values for the EEX market using S_t and $F(t, T_1)$.

	Whole	Part 1	Part 2
	Estimator	Estimator	Estimator
λ_X	98.507	116.807	98.227
λ_L	0.441	1.231	0.579
σ_X	3.161	3.191	2.965
σ_L	0.537	0.337	0.507
L	3.169	3.168	3.222
ρ	0.021	0.019	0.114
σ_S	0.000	0.001	0.047
σ_{Z_1}	0.089	0.074	0.065
σ_{Z_2}	0.000	0.006	0.024
CPU time	1440.58	1208.56	703.68

Table 3.7: Estimated values for the EEX market using S_t , $F(t, T_1)$ and $F(t, T_2)$.

this gives an standard deviation of 0.57. The corresponding quantity for the X_t process is 0.23, so both components contribute significantly to the long run variance of the log-spot price.

In order to investigate if the estimated parameters make sense, we simulated a path of the spot price, future prices and long-term mean process that describe the model using the estimated values from Table 3.6 Part 1, see Figure 3.6.

We now compare those simulations with Figure 3.5, which gives the real

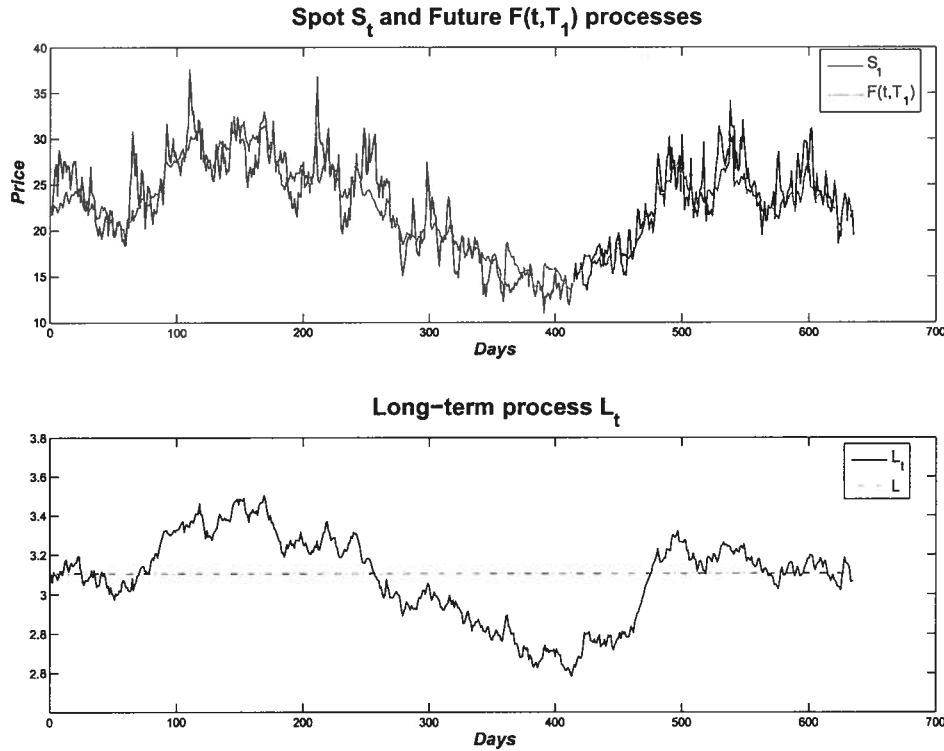


Figure 3.6: Simulation of spot and future prices (upper graph) and long-term mean process (lower graph) using estimate values for part 1.

deseasonalized data. The generated trajectories differ somewhat from those observed in the EEX market. The most notable difference is the absence of price spikes: the simulated data is all in the range €10 - €40, while the real deseasonalized data has about 6 spikes with prices above €70. The absence of such spikes is not surprising since the model contains no mechanism for generating them.

The second feature is that, as with the real data, the spot and futures price do tend to follow each other. This is not surprising, since both prices do relate to the same commodity. Since the long run process L_t is available for the simulated data, we have also plotted this - see lower graph in Figure 3.6. Comparing this with the future price in upper graph we noted that for those parameters, the dominant effect on the future price is from the oscillations

of the long-term mean.

The graphs in Figure 3.6 and Table 3.7 together show that while this model does capture some features of the real data, it does have significant defects, in that it does a poor job of capturing the extreme events, spikes or jumps, which appear in the real market.

Another test for the appropriateness of the model is to compare empirical moments for the real data with those from simulated data.

If we compare the first four empirical central moments of the log-return⁸ of the sequence of the simulated and real spot prices we can see that there is a good fit for the mean value and for the standard deviation, see Table 3.8. The empirical distribution has fatter tails than the normal distribution (kurtosis > 3), indicating a higher occurrence of extreme events, i.e. jumps.

	Real data (1)	Sim. data (1)	Real data (2)	Sim. data (2)
Mean	-0.0019	-0.0002	0.0000	0.0002
Std Dev.	0.2214	0.1770	0.2066	0.20402
Skewness	-1.3125	-0.0037	-0.0800	-0.0160
Kurtosis	29.8250	2.9855	8.1288	2.9313
n data	634	634	631	631

Table 3.8: The table shows the first four moments of the logarithmic de-seasonalized price returns of observed data and the average of 50 simulated trajectories.

⁸Log-return for a sequence of prices S_i are defined as $r_i = \ln(S_{i+1}/S_i)$

Chapter 4

MROU with jumps

In this chapter we consider a jump-diffusion model. This model is similar to the ones considered by Hikspoors and Jaimungal [44] and Nomikos and Soldatos [64]. [44] considers a model of the form

$$\tilde{S}_t = \exp\{h(t) + X_t + J_t\}, \quad (4.1)$$

where X_t is the first component of a pair (X_t, Y_t) satisfying (1.9), and J_t is an independent jump process, see equation (1.11).

In one respect our model represents a simplification of the model in [44], in that there is only one Gaussian factor. However, unlike the model given by (4.1), the Gaussian and jump component in our model are not easily separated into independent processes.

There are several reasons for considering a jump-diffusion model. First, actual spot prices do exhibit spikes - see Figure 3.3, and adding jumps to the process is one way of modeling this. Second, data for spot prices show that the fourth central moment (kurtosis) of the log-returns is much bigger than 3 (see Table 3.8). Diffusion models such as the MROU model in Chapter 3 tend to give the kurtosis close to 3.

4.1 Description of the model

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{Q})$ be a complete filtered probability space. The dynamics of the state variables are given by the following stochastic differential equations:

$$\tilde{S}_t = e^{X_t + h(t)}, \quad (4.2)$$

$$dX_t = -\lambda_X(X_t - L_t) dt + \sigma_X dW_t^1 + J_t^u dN_t^u - J_t^d dN_t^d, \quad (4.3)$$

$$dL_t = -\lambda_L(L_t - L) dt + \sigma_L dW_t^2. \quad (4.4)$$

As before \tilde{S}_t is the spot price. Here the Brownian motions W_t^1 and W_t^2 are independent. The jump behavior of X_t is governed by two types of jumps: upward jumps and downward jumps. The upward jumps J_t^u are exponentially distributed with positive mean $1/\eta_u$, and the downward jumps J_t^d are also exponentially distributed with mean $1/\eta_d$. In this model, N_t^u and N_t^d are two independent Poisson processes with arrival rates λ_u and λ_d respectively. The function $h(t)$ denotes a deterministic seasonality function.

4.2 Valuation of electricity futures

This model also belongs to the class of (AJD) process. We can rewrite equations (4.3) and (4.4) according to equation (A-1) in Appendix A to obtain:

$$\begin{aligned} d \begin{bmatrix} X_t \\ L_t \end{bmatrix} &= \begin{bmatrix} 0 \\ \lambda_L L \end{bmatrix} dt + \begin{bmatrix} -\lambda_X & \lambda_X \\ 0 & -\lambda_L \end{bmatrix} \begin{bmatrix} X_t \\ L_t \end{bmatrix} dt + \begin{bmatrix} \sigma_X & 0 \\ 0 & \sigma_L \end{bmatrix} \begin{bmatrix} dW_t^1 \\ dW_t^2 \end{bmatrix} \\ &+ \begin{bmatrix} J_t^u dN_t^u - J_t^d dN_t^d \\ 0 \end{bmatrix}. \end{aligned}$$

Again defining $\mathbf{U}_t = (X_t, L_t)'$, the expressions for

$$\mathbf{K}_0, \mathbf{K}_1, \mathbf{H}_0, \text{ and } \mathbf{H}_1$$

remain the same as for the MROU, see equation (3.8). However we now have

$$l_0 = \lambda_u + \lambda_d. \quad (4.5)$$

To obtain the formula for the future contracts $F(t, T)$ we need to calculate the *jump transform function* in order to include them in the ODE for $M(t, T)$. The other two equations $N_1(t, T)$ and $N_2(t, T)$ are the same: see equations (3.16) and (3.17).

The density of the distribution of jumps of X_t is given by

$$v_X(x) = \begin{cases} \frac{\lambda_u}{\lambda_u + \lambda_d} e^{-\eta_u x}, & x > 0, \\ \frac{\lambda_d}{\lambda_u + \lambda_d} e^{-\eta_d |x|}, & x < 0. \end{cases}$$

The transform for the jumps is given by (A-4),

$$\varphi(N_1, N_2) = \int_{\mathbb{R}^2} e^{N_1 z_1 + N_2 z_2} d\varphi(z_1, z_2), \quad (4.6)$$

where $\varphi(\cdot)$ is the jump distribution on \mathbb{R}^2 , and N_1, N_2 are such that the integral (4.6) converges. However, since L_t does not have jumps, φ is concentrated on the subspace $z_2 = 0$ and we have

$$\varphi(N_1, N_2) = \int_{\mathbb{R}} e^{N_1 z} d\varphi(z), \quad (4.7)$$

wherever this integral converges, that is, wherever $-\eta_d < N_1 < \eta_u$. Then

$$\begin{aligned} \varphi(N_1, N_2) &= \int_{-\infty}^{\infty} \exp\{N_1 z\} v(z) dz \\ &= \frac{\lambda_u}{\lambda_u + \lambda_d} \int_0^{\infty} \eta_u e^{-(\eta_u - N_1)z} dz + \frac{\lambda_d}{\lambda_u + \lambda_d} \int_{-\infty}^0 \eta_d e^{(\eta_d + N_1)z} dz \\ &= \frac{\lambda_u}{\lambda_u + \lambda_d} \left(\frac{\eta_u}{\eta_u - N_1} \right) + \frac{\lambda_d}{\lambda_u + \lambda_d} \left(\frac{\eta_d}{\eta_d + N_1} \right). \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial M}{\partial t} &= -\lambda_L L N_2 - \frac{1}{2} (\sigma_X^2 N_1^2 + \sigma_L^2 N_2^2) \\ &\quad - \lambda_u \left(\frac{\eta_u}{\eta_u - N_1(t, T)} - 1 \right) - \lambda_d \left(\frac{\eta_d}{\eta_d + N_1(t, T)} - 1 \right), \end{aligned}$$

with the boundary condition $M((1, 0)', T, T) = 0$.

The equations for $N(t, T)$ are, as before,

$$\frac{\partial N_1}{\partial t} = \lambda_X N_1, \quad \text{and} \quad \frac{\partial N_2}{\partial t} = -\lambda_X N_1 + \lambda_L N_2.$$

We can solve the system with the corresponding boundary conditions to obtain

$$\begin{aligned} M(t, T) = & m_1(e^{\lambda_X(t-T)} - 1) + m_2(e^{\lambda_L(t-T)} - 1) + m_3(e^{2\lambda_X(t-T)} - 1) \\ & + m_4(e^{(\lambda_X + \lambda_L)(t-T)} - 1) + m_5(e^{2\lambda_L(t-T)} - 1) \\ & + \frac{\lambda_u}{\lambda_X} \ln \left(\frac{\eta_u - e^{\lambda_X(t-T)}}{\eta_u - 1} \right) + \frac{\lambda_d}{\lambda_X} \ln \left(\frac{\eta_d + e^{\lambda_X(t-T)}}{\eta_d + 1} \right), \end{aligned} \quad (4.8)$$

where the constants m_1, m_2, \dots, m_5 and the solution for $N_1(t, T)$ and $N_2(t, T)$ are given by equations (3.16) and (3.17). See Appendix B for the calculation of $M(t, T)$.

Thus, the expression for the future prices is given by

$$\tilde{F}(t, T) = \exp\{h(T)\} \exp\{M(t, T) + N_1(t, T)X_t + N_2(t, T)L_t\}, \quad (4.9)$$

where the functions $M(t, T)$, $N_1(t, T)$ and $N_2(t, T)$ are given by equations (3.16), (3.17), and (4.8).

Note that (4.8) requires that $\eta_u < 1$. This restriction is to be expected, since the future price is given by

$$\tilde{F}(t, T) = E(\tilde{S}_T | \mathcal{F}_t) = E(e^{X_T + h(T)} | \mathcal{F}_t). \quad (4.10)$$

Since X_T contains, in general, terms involving jumps with an exponential distribution, the expectation in (4.10) will diverge if the upper tail of the jump distribution is sufficiently large. See [44].

4.3 Particle filter setup

Since this model is non-Gaussian, due to the jump process $J_t^u dN_t^u - J_t^d dN_t^d$, we cannot employ the Kalman filter to estimate the parameters λ_u , λ_d , η_u , η_d . We therefore wish to employ the particle filter, which is in principle able to handle quite general distributions.

As for the MROU model, we assume the data is given by

$$X_t^*, \quad Z_{t,T_1}^*, \quad t = 1, 2, \dots, n.$$

(For simplicity we just considered one future price). Here X_t^* is the deseasonalized log-spot price, and Z_{t,T_1}^* the deseasonalized log-future price $F(t, T_1)$.

The first step is discretise the model. Let t_0, t_1, \dots, t_n be the times, and $\Delta t = t_i - t_{i-1}$. As before for simplicity we abuse notation slightly, and write X_t for X_{t_i} . Then we have, using the forward Euler approximation:

$$\begin{aligned} X_t &= X_{t-1} - \lambda_X(X_{t-1} - L_{t-1})\Delta t + \sigma_X \Delta W_t^1 + J_t^u \Delta N_t^u - J_t^d \Delta N_t^d, \\ L_t &= L_{t-1} - \lambda_L(L_{t-1} - L)\Delta t + \sigma_L \Delta W_t^2. \end{aligned}$$

Here J_t^u, J_t^d are exponential random variables with parameters η_u, η_d respectively and

$$\Delta W_t^1 \sim N(0, \Delta t \sigma_X^2), \quad \Delta W_t^2 \sim N(0, \Delta t \sigma_L^2).$$

$\Delta N_t^u, \Delta N_t^d$ are Poisson random variables with parameters $\lambda_u \Delta t, \lambda_d \Delta t$ respectively. Since

$$P(\Delta N_t^u \geq 2) = 1 - e^{-\lambda_u \Delta t} - \lambda_u \Delta t e^{-\lambda_u \Delta t} = O((\Delta t)^2),$$

the probability of two or more jumps in one day is small. We therefore approximate ΔN_t^u by a Bernoulli random variable with parameter $\lambda_u \Delta t$, that is we take

$$P(\Delta N_t^u = 1) = \lambda_u \Delta t, \quad P(\Delta N_t^u = 0) = 1 - \lambda_u \Delta t,$$

with similar approximation for ΔN_t^d .

The transition equation is therefore

$$\begin{aligned} u_t = \begin{pmatrix} X_t \\ L_t \end{pmatrix} &= \begin{pmatrix} 0 \\ \lambda_L L \Delta t \end{pmatrix} + \begin{pmatrix} 1 - \lambda_X \Delta t & \lambda_X \Delta t \\ 0 & 1 - \lambda_L \Delta t \end{pmatrix} \begin{pmatrix} X_{t-1} \\ L_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} \sigma_X \sqrt{\Delta t} W_t^1 + J_t^u \Delta N_t^u - J_t^d \Delta N_t^d \\ \sigma_L \sqrt{\Delta t} \xi_t^0 \end{pmatrix}. \end{aligned} \quad (4.11)$$

As before, we assume the measurements are subject to noise, so the measurement equation is

$$\begin{aligned} v_t = \begin{pmatrix} X_t^* \\ Z_{t,T_i} \end{pmatrix} &= \begin{pmatrix} 0 \\ M(t, T_i) \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ N_1(t, T_i) & N_2(t, T_i) \end{pmatrix} \begin{pmatrix} X_t \\ L_t \end{pmatrix} \\ &+ \begin{pmatrix} \sigma_S \sqrt{\Delta t} & 0 \\ 0 & \sigma_{Z_i} \sqrt{\Delta t} \end{pmatrix} \begin{pmatrix} \xi_t^1 \\ \xi_t^2 \end{pmatrix} \end{aligned} \quad (4.12)$$

where ξ_t^0 , ξ_t^1 , and ξ_t^2 are standard normal random variables.

For the particle filter we initially assume the parameter θ is known, and set it up by writing in the following way. We write $u_t^i(k)$ for the k th component ($k = 1, 2$) of the i th particle ($i = 1, 2, \dots, n_p$) at time t .

A key part of the implementation of the particle filter is the choice of the proposal density $q(u_t|u_{t-1}, v_{1:t})$. One choice would be to simply take $q(\cdot)$ to be the transition density $p(u_t|u_{t-1})$ arising from (4.11). However, this choice is not likely to be optimal. Most of the time, we will have $\Delta N_t^u = \Delta N_t^d = 0$, so most particles will not make a jump at time t . If the observed process X_t^* does make a jump at time t , then most particles will be left behind by this jump. It is therefore better to choose $q(u_t|u_{t-1}, v_{1:t})$ to exploit the information available in v_t . (We remark that the optimal choice of $q(u_t|u_{t-1}, v_{1:t})$ would be the distribution $p(u_t|u_{t-1}, v_{1:t})$, but it is not feasible to calculate this distribution).

We therefore choose $q(\cdot)$ so that the particles are propagated according to the following equation:

$$u_t^{(i)}(1) = X_t^* + \sigma_S \Delta t \xi_t^1, \quad (4.13)$$

$$u_t^{(i)}(2) = -\lambda_L(u_{t-1}^{(i)}(1) - L)\Delta t + \sigma_L \Delta t \xi_t^2. \quad (4.14)$$

Note this means that the second component of $u_t^{(i)}$ just follows the evolution equation of L_t , but we use the new data available in X_t^* to move the first coordinate of the particles close to X_t^* . Here ξ_t^i are independent $N(0, 1)$ random variables ($i = 1, 2$).

We now calculate the likelihood and prior densities. We have

$$\begin{aligned} p(v_t | u_t^{(i)}) &= p(X_t^*, Z_{t,T} | u_t^{(i)}(1), u_t^{(i)}(2)) \\ &= p(X_t^* | u_t^{(i)}(1)) p(Z_{t,T} | u_t^{(i)}(2)), \end{aligned} \quad (4.15)$$

where

$$X_t^* | u_t^{(i)}(1) \sim N(u_t^{(i)}(1), \sigma_S^2) \quad \text{and} \quad Z_{t,T} | u_t^{(i)}(2) \sim N(m_Z, \sigma_Z^2) \quad (4.16)$$

with $m_Z = M(t, T) + N_1(t, T)u_t^{(i)}(1) + N_1(t, T)u_t^{(i)}(2)$.

Now, for the prior density or transition density,

$$\begin{aligned} p(u_t^{(i)} | u_{t-1}^{(i)}) &= p(u_t^{(i)}(1), u_t^{(i)}(2) | u_{t-1}^{(i)}(1), u_{t-1}^{(i)}(2)) \\ &= p(u_t^{(i)}(1) | u_{t-1}^{(i)}(1)) p(u_t^{(i)}(2) | u_{t-1}^{(i)}(2)). \end{aligned} \quad (4.17)$$

Here

$$u_t^{(i)}(2) | u_{t-1}^{(i)}(2) \sim N(m_L, S_L) \quad (4.18)$$

with $m_L = L(1 - e^{-\lambda_L \Delta t}) + u_{t-1}^{(i)}(2)e^{-\lambda_L \Delta t}$ and $S_L = (\sigma_L^2 / 2\lambda_L)(1 - e^{-2\lambda_L \Delta t})$, since L_t is an Ornstein-Uhlenbeck process.

Using the Bernoulli approximation for ΔN_t^u and ΔN_t^d , and neglecting the $O((\Delta t)^2)$ probability of both an upward and downward jump in the same period,

$$\begin{aligned} p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) &= (1 - \lambda_u \Delta t - \lambda_d \Delta t) f_0(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) \\ &\quad + \lambda_u \Delta t f_u(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) + \lambda_d \Delta t f_d(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)). \end{aligned} \quad (4.19)$$

Here f_0 denotes the density corresponding to the jump-free case and f_u and f_d correspond to the case of a single upward and downward jump respectively. Set

$$\mu_X = u_{t-1}^{(i)}(1) - \lambda_X(u_{t-1}^{(i)}(1) - L\lambda_X)\Delta t, \quad \text{and} \quad s^2 = \sigma_X^2 \Delta t,$$

then

$$f_0(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) = (2\pi s^2)^{-1/2} \exp\{-(u_t^{(i)}(1) - \mu_X)^2 / 2s^2\}. \quad (4.20)$$

f_u and f_d are obtained by the convolution of f_0 with the distributions of J^u and J^d . So

$$f_u(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) = \int_0^\infty (2\pi s^2)^{-1/2} e^{-(u_t^{(i)}(1) - y - \mu_X)^2 / 2s^2} \eta_u e^{-\eta_u y} dy. \quad (4.21)$$

Now

$$\begin{aligned}
(u_t^{(i)}(1) - y - \mu_X)^2 / 2s^2 + \eta_u y &= (2s^2)^{-1} \left((y - (u_t^{(i)}(1) - \mu_X))^2 + 2s^2 \eta_u y \right) \\
&= (2s^2)^{-1} \left((y - (u_t^{(i)}(1) - \mu_X))^2 + 2s^2 \eta_u y \right) \\
&= (2b^2)^{-1} \left((y - (u_t^{(i)}(1) - \mu_X - s^2 \eta_u))^2 \right. \\
&\quad \left. - (s^4 \eta_u^2 - 2(u_t^{(i)}(1) - \mu_X) s^2 \eta_u) \right) \\
&= \frac{(y - (u_t^{(i)}(1) - \mu_X - s^2 \eta_u))^2}{2s^2} \\
&\quad - \frac{1}{2} s^2 \eta_u^2 + (u_t^{(i)}(1) - \mu_X) \eta_u.
\end{aligned}$$

Hence writing $A(x) = (u_t^{(i)}(1) - \mu_X - s^2 \eta_u)$,

$$\begin{aligned}
f_u(u_t^{(i)}(1) | u_{t-1}^{(i)}(1)) &= \eta_u e^{s^2 \eta_u^2 / 2} e^{-(u_t^{(i)}(1) - \mu_X) \eta_u} \\
&\quad \int_0^\infty (2\pi s^2)^{-1/2} e^{-(y - (u_t^{(i)}(1) - \mu_X - s^2 \eta_u))^2 / 2s^2} dy \\
&= \eta_u e^{s^2 \eta_u^2 / 2} e^{-(u_t^{(i)}(1) - \mu_X) \eta_u} \int_{-A(x)/s}^\infty (2\pi)^{-1/2} e^{-t^2/2} dt \\
&= \eta e^{s^2 \eta_u^2 / 2} e^{-(u_t^{(i)}(1) - \mu_X) \eta_u} \int_{-\infty}^{A(x)/s} (2\pi)^{-1/2} e^{-t^2/2} dt \\
&= \eta_u e^{s^2 \eta_u^2 / 2} e^{-(u_t^{(i)}(1) - \mu_X) \eta_u} \Phi(A(x)/s),
\end{aligned}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function. So

$$\begin{aligned}
f_u(u_t^{(i)}(1) | u_{t-1}^{(i)}(1)) &= \eta_u e^{s^2 \eta_u^2 / 2} e^{-(u_t^{(i)}(1) - \mu_X) \eta_u} \\
&\quad \Phi((u_t^{(i)}(1) - \mu_X - s^2 \eta_u) / (\sigma_X(\Delta t)^{1/2})).
\end{aligned} \tag{4.22}$$

Let $u' = -u_t^{(i)}(1)$. To get the distribution f_d we use the calculations for f_u ,

$$\begin{aligned}
 f_d(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) &= \int_0^\infty (2\pi s^2)^{-1/2} e^{-(u_t^{(i)}(1)+y-\mu_X)^2/2s^2} \eta_d e^{-\eta_d y} dy \\
 &= \int_0^\infty (2\pi s^2)^{-1/2} e^{-(u'-y-\mu_X)^2/2s^2} \eta_d e^{-\eta_d y} dy \\
 &= \eta_d e^{s^2 \eta_d^2/2} e^{-(u'-\mu_X)\eta_d} \Phi(A(x')/s).
 \end{aligned}$$

So

$$\begin{aligned}
 f_d(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) &= \eta_d e^{s^2 \eta_d^2/2} e^{(u_t^{(i)}(1)+\mu_X)\eta_d} \\
 &\quad \Phi((-u_t^{(i)}(1) - \mu_X - s^2 \eta_d)/(\sigma_X(\Delta t)^{1/2})).
 \end{aligned} \tag{4.23}$$

Combining (4.19), (4.20), (4.22), and (4.23) gives $p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1))$.

Finally, the proposal density is

$$\begin{aligned}
 q(u_t^{(i)}|u_{t-1}^{(i)}, v_t) &= p(u_t^{(i)}(1), u_t^{(i)}(2)|u_{t-1}^{(i)}(1), u_{t-1}^{(i)}(2), v_t) \\
 &= p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1), v_t) p(u_t^{(i)}(2)|u_{t-1}^{(i)}(2)),
 \end{aligned} \tag{4.24}$$

with $p(u_t^{(i)}(2)|u_{t-1}^{(i)}(2))$ given as (4.18) and

$$p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1), v_t) = (2\pi\sigma_S^2)^{-1/2} \exp\{-(u_t^{(i)}(1) - X_t^*)^2/2\sigma_S^2\}. \tag{4.25}$$

Combining (4.15), (4.17), and (4.24) we calculate the associated weights

$$\begin{aligned}
 w_t^{(i)} &= \frac{p(v_t|u_t^{(i)}) p(u_t^{(i)}|u_{t-1}^{(i)})}{q(u_t^{(i)}|u_{t-1}^{(i)}, v_t)} \\
 &= \frac{p(X_t^*|u_t^{(i)}(1)) p(Z_{t,T}|u_t^{(i)}(2)) p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)) p(u_t^{(i)}(2)|u_{t-1}^{(i)}(2))}{p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1), v_t) p(u_t^{(i)}(2)|u_{t-1}^{(i)}(2))} \\
 &= p(Z_{t,T}|u_t^{(i)}(2)) p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1)).
 \end{aligned}$$

Here the first term is given by (4.16), and the second by combining (4.19), (4.20), (4.22), and (4.23).

4.4 Simulated data with known parameters

We tested the implementation of the particle filter presented in Section (4.3) for the MROU jump model to estimate the long-term mean process L_t . We simulated a series of 100 time data points according to (4.3), (4.4), and (4.9), taking $\lambda_X = 110$, $\lambda_L = 5$, $\sigma_X = 3$, $\sigma_L = 1$, $L = 3.2$, $\eta_u = 1.5$, $\eta_d = 3$, $\lambda_u = 5$, and $\lambda_d = 1$. Figure 4.1 shows the particle filter estimates of the state using 100 particles.

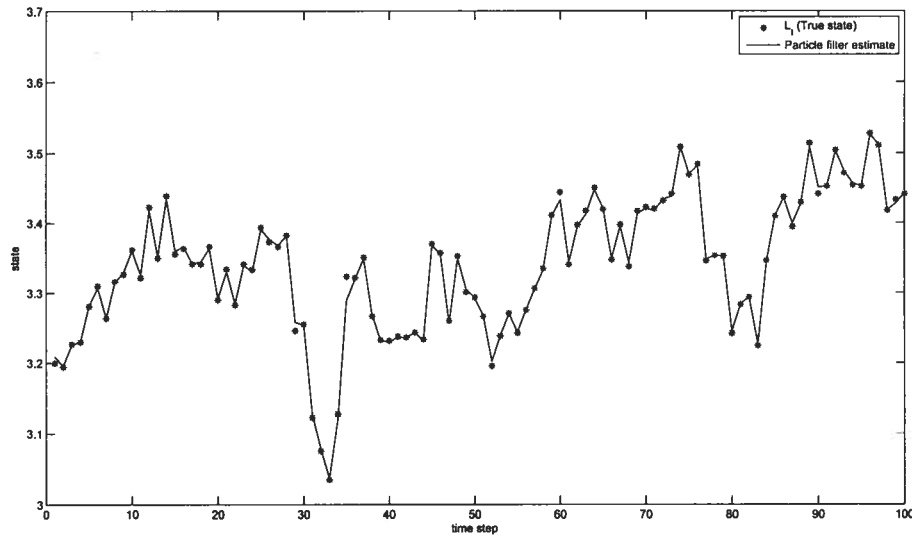


Figure 4.1: Plot of the true state L_t and estimate of the particle filter.

As is clear from Figure 4.1 the particle filter does a good job of estimating the transition state for this example even using just a few particles, provided the parameters are known.

4.5 Likelihood function estimation

Using the particle filter for a fixed parameter θ , one can obtain an estimate of the likelihood function $\mathcal{L}_\theta(v_{1:t})$ by (2.45). However, as we already mentioned above, severe difficulties arise when one tries to optimize this function since, for each value of θ one is using a different randomization.

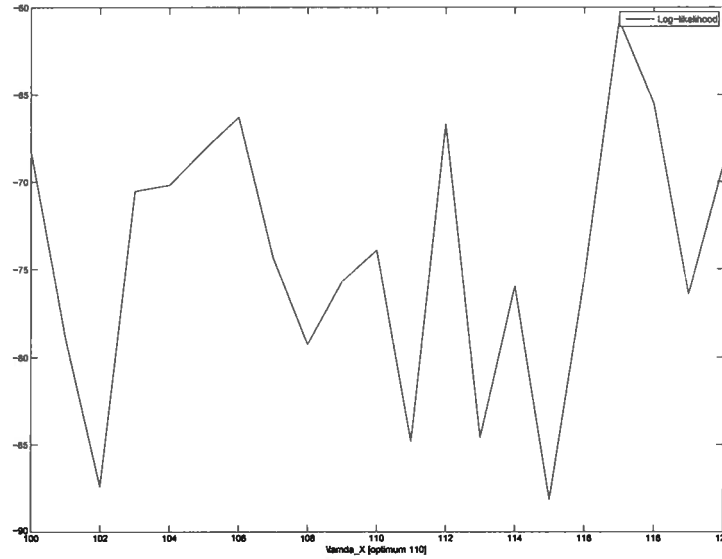


Figure 4.2: The log-likelihood for different λ_X values.

To investigate the problem, we fixed all the parameters except λ_X at their correct value ($\lambda_X = 110$), and estimated $\mathcal{L}_{\lambda_X}(v_{1:t})$ for simulated data. Taking 1000 points and 2000 particles, we estimated $\mathcal{L}_{\lambda_X}(v_{1:t})$ for λ_X in integer steps between 102 and 120. As is clear from Figure 4.2, there is little prospect of satisfactory use of a hill-climbing algorithm given the level of noise. We repeated this estimation with 15000 particles and obtained a similar curve, but with oscillations roughly 10 times smaller. These oscillations however were still large enough so that a hill-climbing algorithm would have difficulties in locating the optimum.

It seems likely that increasing the number of particles by a factor of k will increase accuracy of the estimates of the log-likelihood by \sqrt{k} . If so, one

might need 10^6 - 10^8 particles in order to use a hill-climbing approach even in one dimension.

4.6 Sequential parameters

Our goal is to estimate the long-term mean process together with the 11-dimensional parameter vector θ given by

$$\theta = \{\lambda_X, \lambda_L, \sigma_X, \sigma_L, \sigma_S, \sigma_Z, L, \eta_u, \eta_d, \lambda_u, \lambda_d\},$$

according to the information available at a given time t , i.e. $v_{1:t}$. As we explained in Section 2.6, from a Bayesian point of view we concatenate the state vector and the parameters and apply the filter to this augmented state. Then we define

$$y_t = \begin{pmatrix} u_t \\ \theta_t \end{pmatrix} \quad (4.26)$$

where θ_t is the particle approximation to θ . It is useful to add some noise to the transition for θ_t ⁹

$$\theta_t = \theta_{t-1} + \xi_{t-1} \quad (4.27)$$

where $\{\xi_{t-1}\}_{t \geq 1}$ is a small artificial noise with a decreasing variance Λ_t with t .

A wide variety of choices of ξ_{t-1} , and Λ_t are possible. It is clear that some care has to be taken in the choice of Λ_t . If Λ_t is too large, then particles $\theta_t^{(i)}$ will oscillate too much to give a satisfactory estimate of θ (the larger the covariance, the more quickly older data are discarded). If Λ_t is too small, then unless the initial value is nearly correct, θ_t will not move enough to reach the correct value.

The literature contains a number of suggestions on how Λ_t should be chosen. In [43] the author suggested ξ_{t-1} to be white noise. Others proposed

⁹If not, we only sample particles in Θ space at time 1 and never modify their locations, then after a few time steps $p(\theta|v_{1:t})$ is approximated by a single particle.

to set Λ_t as a diagonal matrix annealing to zero - see [41, 53, 68] for more options.

First we take a fairly simple approach. We will take Λ_t to be diagonal, and write $\theta_{t,j}$ for the j -th component of θ_t and $\phi_{t,j}$ for the j -th element in the diagonal of Λ_t . We set

$$\phi_{t,j} = b_j \left(\frac{c_j}{t + c_j} \right) \quad (4.28)$$

where b_j and c_j are constants. Note that $\sum_{t=1}^{\infty} \text{var}(\phi_{t,j}) = \infty$. The ‘annealing’ given by (4.28) is not fast enough to make $\theta_{t,j}$ converge almost surely.

Later we use the Liu and West [57] approach. They suggest approximating the distribution $p(\theta|v_{1:t})$ by using a mixture of Gaussian distributions, that is

$$p(\theta|v_{1:t}) \approx \sum_{i=1}^{n_p} \tilde{w}_t^{(i)} n(\theta|m_t^i, h^2 \Lambda_t). \quad (4.29)$$

The quantity $m_t^{(i)} = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$ is the kernel location for the i -th component of the mixture where

$$\bar{\theta}_t = \sum_{i=1}^{n_p} \tilde{w}_t^{(i)} \theta_t^{(i)}, \quad (4.30)$$

and the matrix Λ_t is an estimate of the posterior variance covariance-matrix, i.e.

$$\Lambda_t = \sum_{i=1}^{n_p} \tilde{w}_t^{(i)} (\theta_t - \bar{\theta}_t)(\theta_t - \bar{\theta}_t)'. \quad (4.31)$$

The constants h and a , that measure the extent of the shrinkage and the degree of over-dispersion of the mixture, are given by $h^2 = 1 - ((2\varphi - 1)/2\varphi)^2$, $a = \sqrt{1 - h^2}$, where the discount factor φ ranges between 0.95 - 0.99. In this case the artificial dynamic of the parameter is given by

$$\theta_{t+1} = m_t^{(i)} + h\Lambda_t^{1/2} \epsilon_{t+1}, \quad \epsilon_{t+1} \sim n(0, \mathbf{I}_{n_\theta \times n_\theta}), \quad (4.32)$$

where $n(x|\mu, \sigma^2)$ denotes the density of the Gaussian distribution. With these specifications the likelihood, prior and proposal densities for $u_t^{(i)}$ now depend on $\theta_t^{(i)}$. So we have

$$p(v_t|y_t^{(i)}) = n(X_t^*|u_t^{(i)}(1), \sigma_S^2) \times \\ n(Z_{t,T}|M_t(\theta_t^{(i)}) + N_{1,t}(\theta_t^{(i)})u_t^{(i)}(1) + N_{2,t}(\theta_t^{(i)})u_t^{(i)}(2), \sigma_Z^2), \quad (4.33)$$

$$p(y_t^{(i)}|y_{t-1}^{(i)}) = p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1), u_{t-1}^{(i)}(2), \theta_{t-1}^{(i)}) \\ \times p(u_t^{(i)}(2)|u_{t-1}^{(i)}(2), \theta_{t-1}^{(i)}) \prod_{j=1}^m n(\theta_{t,j}^{(i)}|\theta_{t,j}^{(i)}, \phi_{t,j}^{(i)}), \quad (4.34)$$

$$q(y_t^{(i)}|y_{t-1}^{(i)}, v_t) = n(u_t^{(i)}(1)|X_t^*, \sigma_S^2) p(u_t^{(i)}(2)|u_{t-1}^{(i)}(2), \theta_{t-1}^{(i)}) \\ \times \prod_{j=1}^m n(\theta_{t,j}^{(i)}|\theta_{t,j}^{(i)}, \phi_{t,j}^{(i)}). \quad (4.35)$$

Substituting (4.33)-(4.35) into the weights formula gives

$$w_t^{(i)} = \frac{p(v_t|u_t^{(i)}) p(u_t^{(i)}|u_{t-1}^{(i)})}{q(u_t^{(i)}|u_{t-1}^{(i)}, v_t)} \\ = n(Z_{t,T}|M_t(\theta_t^{(i)}) + N_{1,t}(\theta_t^{(i)})u_t^{(i)}(1) + N_{2,t}(\theta_t^{(i)})u_t^{(i)}(2), \sigma_Z^2) \\ \times p(u_t^{(i)}(1)|u_{t-1}^{(i)}(1), u_{t-1}^{(i)}(2), \theta_{t-1}^{(i)}). \quad (4.36)$$

The last density in (4.36) is given by (4.17).

4.7 Empirical results

In this section we report some empirical results based on simulated data to show the performance of the method. The true parameters were fixed to be:

$$\begin{aligned} \lambda_X &= 110 & \sigma_X &= 3 & \lambda_L &= 5 & L &= 3.2 & \sigma_L &= 1 \\ \eta_u &= 1.5 & \lambda_u &= 5 & \eta_d &= 3 & \lambda_d &= 1. \end{aligned}$$

For simplicity we took $\sigma_S = 0.1$ and $\sigma_Z = 0.1$ to be fixed and assumed to be known.

Figures 4.3 and 4.4 show a simulation of a MROU with jumps process using the parameter values as above. The upper graph in Figures 4.3 shows the simulated spot price and the future price with maturity of one month. The lower graph is the long-term mean process L_t . Figure 4.4 shows the upward and downward jump processes. For this data set the skewness and kurtosis is -1.8895, and 21.069 respectively.

Considering the first approach (4.28) and using 1000 particles we ran the particle filter to optimize all the parameters, but the filter failed to obtain good results on most of the parameters.

Difficulties of this kind in a high-dimensional situation are not surprising. Even in the case of the Kalman filter, where we optimized a deterministic function, the hill-climbing algorithm sometimes failed to find a point close to the true maximum.

The particle filter replaces the hill-climbing point with a cloud of particles, where the weights $w_t^{(i)}$ are higher when the particles are close to the true value. This should mean that the particle cloud drifts towards the true value of the parameters, but it seems intuitively unlikely that it will perform as well as a hill-climbing algorithm.

A known weakness of optimization algorithms is the following. The higher the number of parameters, the worse the performance of the algorithm. This means that a one-parameter optimization should perform best. To test this, we allowed in turn each of the parameters to vary. Thus in each run we fixed all but one parameter at its correct value, and ran the particle filter with just one $\theta_{t,j}$ varying. We took the number of particles to be 1000, and the number of time steps to be 1200. We chose $\phi_{t,j}$ as in (4.28), with $c_j = 100$.

The Figure 4.5 shows the dynamic under the artificial noise (4.28) of the parameter $\theta_{t,1} := \lambda_X$, for $t = 1, 2, \dots, 1200$ using 1000 particles and its optimal value (110). To test the particle filter, we started the parameter some way away from its correct value. The noise is decreasing quite slowly,

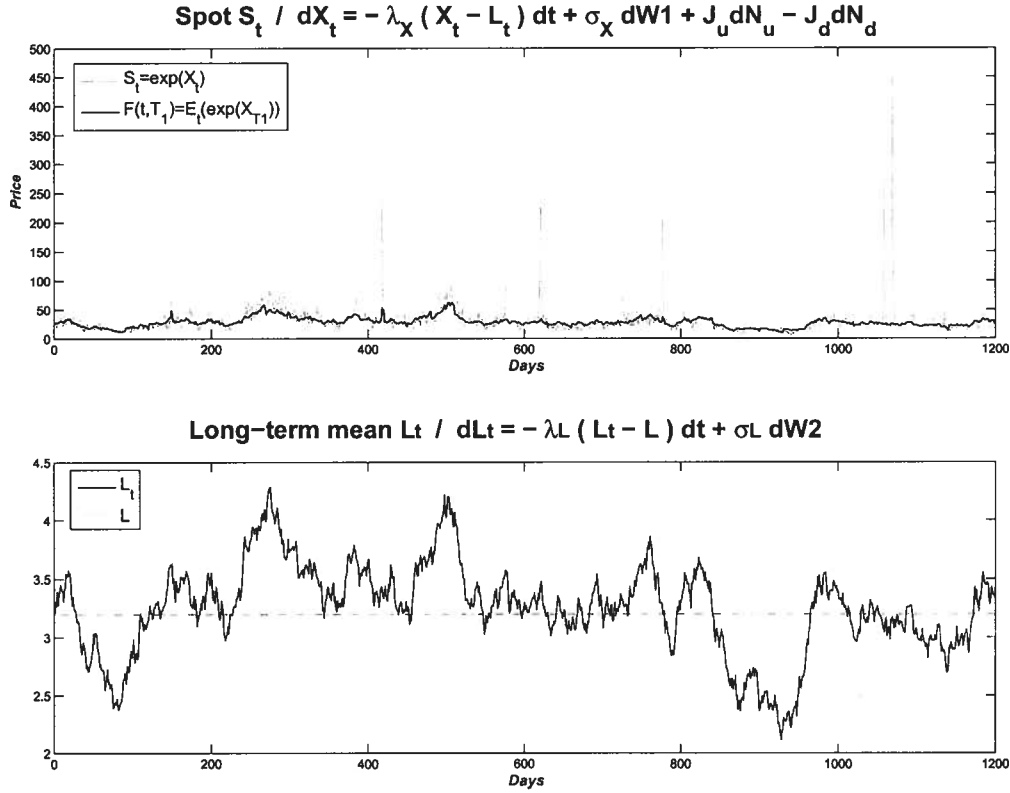


Figure 4.3: The graph shows a simulated trajectory of the MROU model with jumps.

and hence the variance of the process remains large. Figure 4.5 suggests that there is little “push” in the particle filter towards the correct value, and this is confirmed by Table 4.1, which shows a fairly wide dispersion of estimates around the true value, for different runs of the filter on the same data set.

A similar pattern arises for the other parameters.

We ran the particle filter 25 times for each parameter. Let $\theta_{i,j}(k)$ be the value of $\theta_{i,j}$ in the k th run, where $k = 1, 2, \dots, 25$. For each run k , we estimated θ_j by taking its average over the last 400 observations, that is,

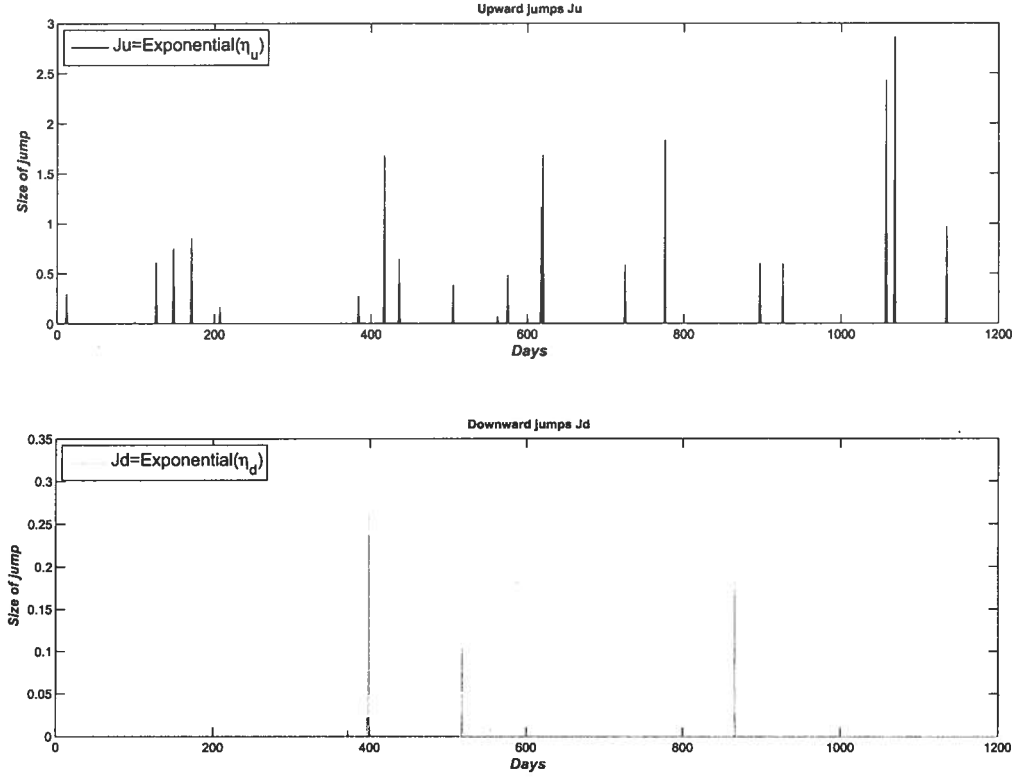


Figure 4.4: The graph shows the upward and downward jumps for the simulated trajectory of the MROU model with jumps.

$$\bar{\theta}_j(k) = \frac{1}{400} \sum_{t=401}^{1200} \theta_{t,j}(k).$$

We then calculated the mean and standard deviation of $\bar{\theta}_j(k)$, $k = 1, 2, \dots, 25$. Table 4.2 summarizes the outcome of the whole procedure. The first column is the true value of the parameters. The second column is the initial value of $\phi_{t,j}$, that is b_j in (4.31). The third and fourth columns give the mean and standard deviation of the estimates ($\bar{\theta}_j(k)$, $k = 1, 2, \dots, 25$).

As we can see from the Table 4.1, the performance results are rather

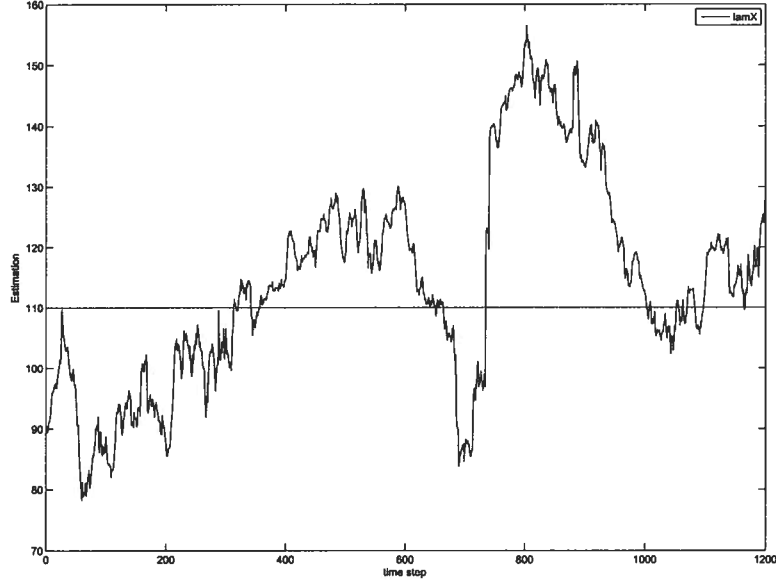


Figure 4.5: Sample particle filter trajectories for the estimate of λ_X .

	True	r1	r2	r3	r4	r5	r6	r7	r8
λ_X	110	126.13	161.49	102.76	87.72	76.41	105.29	124.15	83.71

Table 4.1: Sample of 8 estimated values for λ_X .

	True value	b_j	mean value	std.
λ_X	110	30	99.29	19.64
σ_X	3	1	3.26	0.048
λ_L	5	2	8.147	0.729
σ_L	1	1	1.16	0.017
L	3.2	1	3.37	0.037
η_u	1.5	0.5	2.444	0.754
λ_u	5	0.5	8.301	2.983
η_d	3	1	14.87	6.917
λ_d	1	0.5	4.218	1.817

Table 4.2: Individual estimates for parameters in MROU with jumps model.

mixed. In some cases the particle filter was able to obtain a reasonable estimate for the parameter, but in others, particularly for the parameters associated with the jumps, the estimates are far from the true value. The particle filter literature suggests that even with a very large number of particles one may not be able to obtain accurate results.

Next we tested the Liu and West approach. We used 4000 particles to approximate the distribution of interest. A problem that we noticed is that the estimated posterior variance-covariance matrix Λ_t collapses to zero after a few hundred iterations. We solved this problem choosing an efficient resampling scheme that kept low the variance in the particle filter algorithm. We found that the residual resampling kept the covariance matrix positive.

We ran two examples. In the first one we fixed all but one parameter at its correct value, and ran the algorithm choosing a reasonable initial point for the free parameters. We obtained good results even for the jump parameters. (For simplicity we consider again $\sigma_S = 0.1$ and $\sigma_Z = 0.1$). The results are displayed in Figure 4.6.

It is interesting to note that the algorithm detects precisely the parameters λ_u , λ_d , η_u , η_d associated with extreme events (spikes).

For the second example, we took the vector parameter to estimate as

$$\theta = \{\lambda_X, \lambda_L, \sigma_X, \sigma_L, L\}$$

while the rest of the parameters were fixed to their optimal values. Again, we started the algorithm choosing a reasonable initial point for the free parameters θ as well a small initial variance. Figure 4.7 presents part of the results.

From this example, we noticed that the algorithm implemented gave accurate estimates for the parameters L , σ_L , and σ_X .

More difficulties arose when estimating the speed reversion for long and short-term process λ_X , and λ_L . There are slightly under and over estimated respectively.

Overall, the algorithm provided more precise estimated values but difficulties arose for the parameters related to the jump process. Further, the algorithm required a significant amount of tuning, i.e. choosing the initial

value and variance of the artificial noise. However the graphs in Figures 4.6 - 4.7 indicate a significantly better performance than that obtained by using (4.28). Given these difficulties, we did not feel confident that the algorithm would perform satisfactorily on real data.

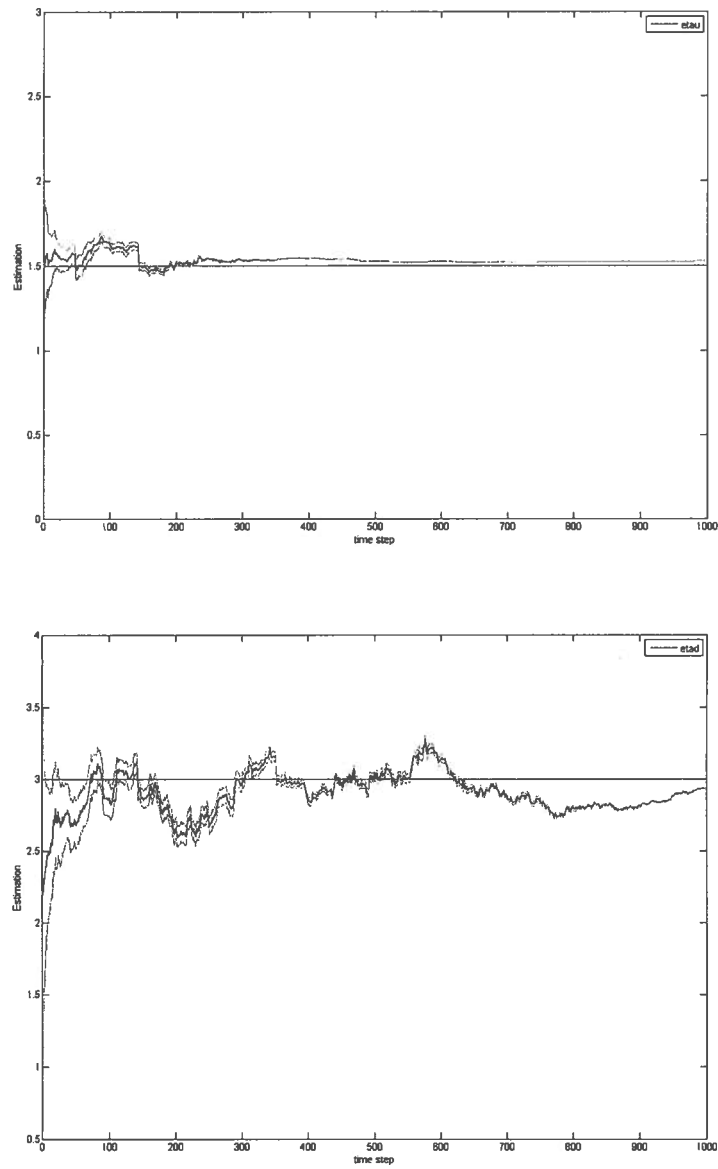


Figure 4.6: Sample particle filter trajectories for the estimate of η_u and η_d using Liu and West approach.

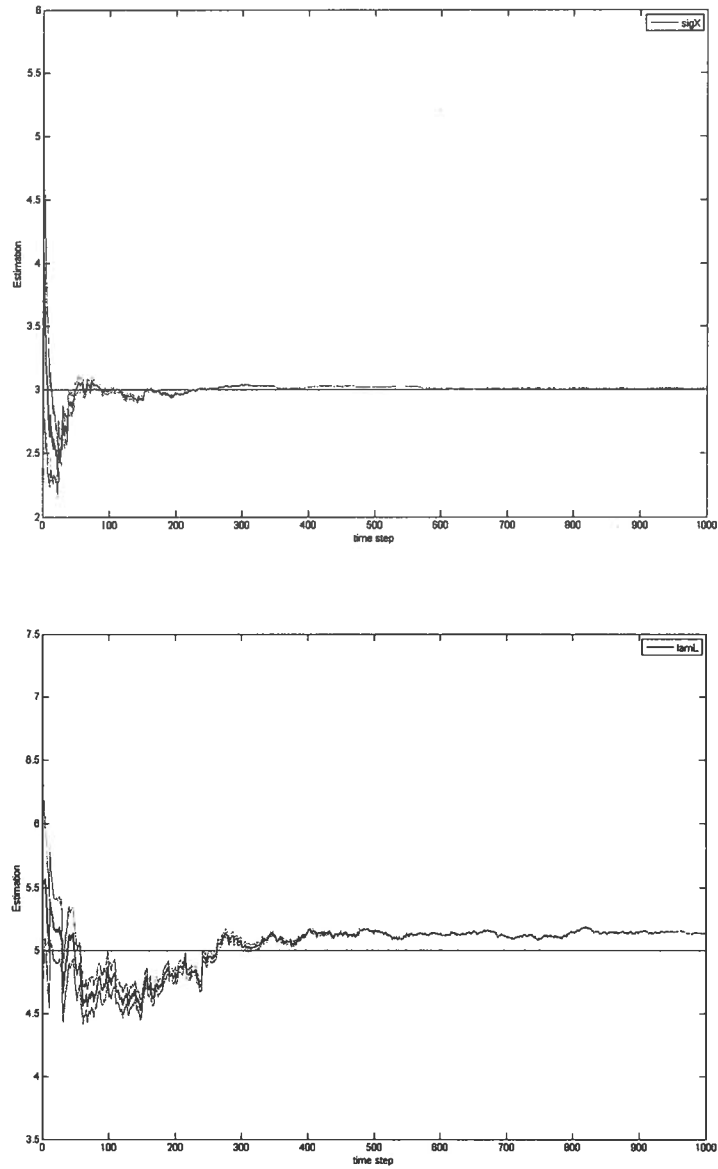


Figure 4.7: Sample particle filter trajectories for the estimate of σ_X and λ_L using Liu and West approach.

Chapter 5

NLMROU model

Our difficulties with the particle filter led us to look for more tractable models which have the potential for explaining price spikes. While the incorporation of jumps is the most natural way to account for price spikes in the spot price, other explanations have been offered. Barlow [3] proposed a non-linear diffusion model, which can produce price spikes similar to those observed in real data. This single factor model is unlikely to provide a good explanation for the observed relation between spot and future prices. Here we present a two-factor model of the same kind. The model is estimated using data from the European Energy Exchange.

5.1 The model

The nonaffine term structure two-factor model for futures prices is known as the Non-Linear Mean-Reversion Ornstein-Uhlenbeck model (NLMROU). It uses the inverse of the Box-Cox transformation to generate price spikes that fit the observed data observed in the power market.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{Q})$ be a filtered probability space. The dynamic of the spot price under the risk-neutral measure \mathbb{Q} is the following:

$$\tilde{S}_t = f_\alpha(X_t)h(t)$$

where X_t is an Ornstein-Uhlenbeck process which reverts to a stochastic mean L_t , also fluctuating as Ornstein-Uhlenbeck processes and f_α is the inverse of the Box-Cox transformation. This transformation was introduced in the context of electricity markets in [3]. The deterministic component $h(t)$ incorporates the seasonality effects in the model.

More precisely,

$$\tilde{S}_t = h(t) \times \begin{cases} (1 + \alpha X_t)^{1/\alpha}, & 1 + \alpha X_t > \varepsilon_0 \\ \varepsilon_0^{1/\alpha}, & 1 + \alpha X_t \leq \varepsilon_0 \end{cases} \quad (5.1)$$

$$dX_t = -\lambda_X(X_t - L_t) dt + \sigma_X dW_t^1, \quad (5.2)$$

$$dL_t = -\lambda_L(L_t - L) dt + \sigma_L dW_t^2. \quad (5.3)$$

where the two Brownian motions W_t^1 , and W_t^2 have correlation ρ and $\alpha \neq 0$. If $\alpha = 0$ then S_t is the MROU model, see Chapter 3, with a cutoff at ε_0 . If $\alpha < 0$, the function $(1 + \alpha X_t)^{1/\alpha}$ increases more rapidly than an exponential function.

An important advantage of this approach compared to other methods for produce spikes in the spot price process is the inclusion of just one more parameter α in the model to be estimated. The deterministic seasonality function is the same as is described by equation (3.33) and its components are estimated by least-square fitting exactly as in Chapter 3. We denote the deseasonalized spot price by S_t , that is

$$S_t = \tilde{S}_t / \hat{h}(t).$$

where as before $\hat{h}(t)$ is the estimated seasonal correction.

5.2 Future price

Based on the risk neutralized process (5.1)-(5.3) we calculate the Future price. Assuming a deterministic interest rate the Future price is the expected future spot price under the risk neutral measure \mathbb{Q} , i.e.

$$\begin{aligned} \tilde{F}(t, T) &= h(T) \mathbb{E}(f_\alpha(X_T) | X_t = x, L_t = l) \\ &= h(T) \int_{-\infty}^{\infty} f_\alpha(w) \frac{1}{\sqrt{2\pi\sigma^2(s)}} \exp \left\{ -\frac{1}{2} \left(\frac{w - \mu(s, x, l)}{\sigma(s)} \right)^2 \right\} dw, \end{aligned} \quad (5.4)$$

where $\mu(s, x, l)$, and $\sigma(s)$ are the mean and variance of X_T respectively.

Taking $y = (w - \mu(s, x, l))/\sigma(s)$

$$\tilde{F}(t, T) = h(T) \int_{-\infty}^{\infty} f_{\alpha}(\mu(s, x, l) + \sigma(s)y) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \quad (5.5)$$

$$\begin{aligned} &= \frac{h(T)}{\sqrt{2\pi}} \int_{-\infty}^{d_1} (1 + \alpha(\mu(s, x, l) + \sigma(s)y))^{1/\alpha} e^{-y^2/2} dy \\ &\quad + A_0 \int_{d_1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \end{aligned} \quad (5.6)$$

$$\begin{aligned} &= \frac{h(T)}{\sqrt{2\pi}} \int_{-\infty}^{d_1} (1 + \alpha(\mu(s, x, l) + \sigma(s)y))^{1/\alpha} e^{-y^2/2} dy \\ &\quad + A_0 \Phi(-d_1), \end{aligned} \quad (5.7)$$

with

$$A_0 = \varepsilon_0^{1/\alpha}, \quad d_1 = \frac{\varepsilon_0 - 1}{\sigma(s)\alpha} - \frac{\mu(s, x, l)}{\sigma(s)},$$

and $\Phi(\cdot)$ denotes the cumulative normal distribution function.

Using the fact that

$$\mathbb{E}(e^Y) = e^{\mu + \sigma^2/2} \quad \text{with } Y \sim N(\mu, \sigma^2)$$

we are able to calculate the mean $\mu(s, x, l)$, and variance $\sigma(s)$ of X_T in (5.4).

Let

$$\alpha_0 = -\frac{\lambda_X}{\lambda_X - \lambda_L}, \quad \text{and} \quad s = T - t.$$

From previous calculations, see equation (3.9) and Appendix (6.1)

$$\mathbb{E}(\exp\{(u_1, u_2) \cdot (X_T, L_T)'\} | (X_t, L_t)'\} = \exp\{M(s) + N_1(s)X_t + N_2(s)L_t\}$$

where

$$N_1(s) = u_1 \theta_X(s), \quad N_2(s) = u_1 \alpha_0 (\theta_X(s) - \theta_L(s)) + u_2 \theta_L(s),$$

and

$$\begin{aligned}
 M(s) = & \frac{-\lambda_L L u_1 \alpha_0}{\lambda_X} (\theta_X(s) - 1) + \left(\frac{\lambda_L L u_1 \alpha_0 - \lambda_L L u_2}{\lambda_L} \right) (\theta_L(s) - 1) \\
 & + \left(\frac{-\sigma_X^2 u_1^2 - \sigma_L^2 u_1^2 \alpha_0^2 - 2\rho\sigma_X\sigma_L u_1^2 \alpha_0}{4\lambda_X} \right) (\theta_{2X}(s) - 1) \\
 & + \left(\frac{-\sigma_L^2 u_1^2 \alpha_0^2 - \sigma_L^2 u_2^2 + 2\sigma_L^2 u_1 u_2 \alpha_0}{4\lambda_L} \right) (\theta_{2L}(s) - 1) \\
 & + \left(\frac{\sigma_L^2 u_1^2 \alpha_0^2 - \sigma_L^2 u_1 u_2 \alpha_0 + \rho\sigma_X\sigma_L u_1^2 \alpha_0 - \rho\sigma_X\sigma_L u_1 u_2}{\lambda_X + \lambda_L} \right) (\theta_{X+L}(s) - 1).
 \end{aligned}$$

Taking $u_2 = 0$

$$\begin{aligned}
 \mathbb{E}(\exp\{u_1 X_T\} | (X_t = x, L_t = l)') = \\
 \exp\{u_1 (M_{11}(s) + N_{11}(s)x + N_{21}(s)l) + \frac{u_1^2}{2} M_{12}(s)\} \quad (5.8)
 \end{aligned}$$

where

$$\begin{aligned}
 M_{11}(s) &= \frac{-\lambda_L L \alpha_0}{\lambda_X} (\theta_X(s) - 1) + \frac{\lambda_L L \alpha_0}{\lambda_L} (\theta_L(s) - 1), \\
 N_{11}(s) &= \theta_X(s), \\
 N_{21}(s) &= \alpha_0 (\theta_X(s) - \theta_L(s)),
 \end{aligned}$$

and

$$\begin{aligned}
 M_{12}(s) &= \frac{-\sigma_X^2 - \sigma_L^2 \alpha_0^2 - 2\rho\sigma_X\sigma_L \alpha_0}{2\lambda_X} (\theta_X^2 - 1) \\
 &\quad - \frac{\sigma_L^2 \alpha_0^2}{2\lambda_L} (\theta_{2L}(s) - 1) + \frac{2\sigma_L^2 \alpha_0^2 + 2\rho\sigma_X\sigma_L \alpha_0}{(\lambda_X + \lambda_L)} (\theta_{X+L}(s) - 1).
 \end{aligned}$$

Therefore the mean and variance of X_T are respectively

$$\mu(s, x, l) = M_{11}(s) + N_{11}(s)x + N_{21}(s)l$$

and

$$\sigma(s) = M_{12}(s).$$

Unfortunately the integral in (5.7) does not admit a closed-form solution. However, this is not a significant obstacle, since these integrals can be evaluated quickly by numerical methods. The Futures prices have been calculated for this model, though in a rather non-explicit fashion - see [78].

5.3 Unscented Kalman filter setup and estimation procedure

This model is non-linear but has Gaussian noise, so an appropriate technique is to use the Extended Kalman filter (EKF) or the Unscented Kalman filter (UKF), see Section 2.3.

The starting point of our inferential procedure involves an Euler discretization. The model is thus evaluated at a set of discrete times $\{t_i : i = 0, 1, \dots, n\}$ such that $\Delta t = t_i - t_{i-1}$. Writing X_t for X_{t_i} as before, the Euler scheme for equations (5.1)-(5.3) can be written as:

$$S_t = f_\alpha(X_t), \quad (5.9)$$

$$X_t = X_{t-1} - \lambda_X(X_{t-1} - L_{t-1})\Delta t + \sqrt{1 - \rho^2}\sigma_X\Delta W_t^1 + \rho\sigma_X\Delta W_t^2, \quad (5.10)$$

$$L_t = L_{t-1} - \lambda_L(L_{t-1} - L)\Delta t + \sigma_L\Delta W_t^2. \quad (5.11)$$

In order to apply the Unscented Kalman filter, we use the state-space representation for the NLMROU model. We defined the state equation as

$$\begin{aligned} u_t = \begin{pmatrix} X_t \\ L_t \end{pmatrix} &= \begin{pmatrix} 0 \\ \lambda_L L \Delta t \end{pmatrix} + \begin{pmatrix} 1 - \lambda_X \Delta t & \lambda_X \Delta t \\ 0 & 1 - \lambda_L \Delta t \end{pmatrix} \begin{pmatrix} X_{t-1} \\ L_{t-1} \end{pmatrix} \\ &+ \begin{pmatrix} \sqrt{1 - \rho^2}\sigma_X\Delta W_t^1 + \rho\sigma_X\Delta W_t^2 \\ \sigma_L\sqrt{\Delta t}W_t^2 \end{pmatrix}. \end{aligned} \quad (5.12)$$

The observations we have available, after seasonal correction, are the spot and future prices $(S_t, F(t, T))$. However, the UKF allows us to take as

our measurement any transformation of these observations. To reduce non linearity in the model we therefore took our measurement equation as

$$\begin{aligned}
 v_t &= \begin{pmatrix} \log S_t \\ \log F(t, T) \end{pmatrix} + \begin{pmatrix} \sigma_S \sqrt{\Delta t} W_t^1 \\ \sigma_Z \sqrt{\Delta t} W_t^2 \end{pmatrix}, \\
 &= \begin{pmatrix} \log f_\alpha(X_t) \\ \log \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_1} (1 + \alpha(\mu(s, x, l) + \sigma(s)y))^{1/\alpha} e^{-y^2/2} dy + A_0 \Phi(-d_1) \right\} \end{pmatrix} \\
 &\quad + \begin{pmatrix} \sigma_S \sqrt{\Delta t} W_t^1 \\ \sigma_Z \sqrt{\Delta t} W_t^2 \end{pmatrix},
 \end{aligned} \tag{5.13}$$

where S_t denotes the deseasonalized spot price, and $F(t, T)$ is the deseasonalized future price at time t with maturity T . In this case the noise covariances are given by

$$\Sigma_q = \text{cov} \begin{pmatrix} \sqrt{1 - \rho^2} \sigma_X \Delta W_t^1 + \rho \sigma_X \Delta W_t^2 \\ \sigma_L \sqrt{\Delta t} W_t^2 \end{pmatrix} = \begin{pmatrix} \sigma_X^2 \Delta t & \rho \sigma_X \sigma_L \Delta t \\ \rho \sigma_X \sigma_L \Delta t & \sigma_L^2 \Delta t \end{pmatrix},$$

and

$$\Sigma_r = \text{cov} \begin{pmatrix} \sigma_S \Delta W_t^1 \\ \sigma_Z \Delta W_t^2 \end{pmatrix} = \begin{pmatrix} \sigma_S^2 & 0 \\ 0 & \sigma_Z^2 \end{pmatrix}.$$

We set up the specific characteristics of the state-space model for the spot and future prices using the transition and measurement equations (5.12) and (5.13). The UKF parameters were set to $\alpha = 0.001$, $\beta = 2$, and $\kappa = 0$. Based on this state-space formulation we are able to run the unscented Kalman filter algorithm in order to estimate the parameter set

$$\theta = \{\lambda_X, \lambda_L, \sigma_X, \sigma_L, L, \rho, \alpha, \sigma_S, \sigma_Z\}$$

by means of maximum likelihood according to Section (2.5). We run Algorithm 4 in Section 3.4, but using the UKF instead of the Kalman filter in (Step 2) to obtain $\hat{\theta}$.

The log-likelihood function is calculated as

$$\mathcal{L}_\theta(v_{1:n}) \propto - \sum_{k=1}^n [\log |\Sigma_{v_k v_k}| + (v_k - \bar{v}_{k|k-1})' \Sigma_{v_k v_k}^{-1} (v_k - \bar{v}_{k|k-1})] \quad (5.14)$$

where $\Sigma_{v_k v_k}$, and $\bar{v}_{k|k-1}$ are given by (2.23) and (2.22).

5.4 Simulation results

Assuming the parameters are known, and using the Euler scheme discretization in (5.9)-(5.11) we simulated a path of the deseasonalized spot and future price with $n = 800$ observations of the NLMROU model using $\lambda_X = 100$, $\lambda_L = 3$, $\sigma_X = 1.7$, $\sigma_L = 0.4$, $L = 1.9$, $\rho = 0$, $\alpha = -0.4$ and $\varepsilon_0 = 800^\alpha$ with $\Delta t = 1/250$. Figure 5.1 shows the trajectories.

As we can see in Figure 5.2 the UKF is able to recover the ‘hidden’ states.

For simplicity we fixed the parameters $\rho = 0$, $\sigma_S = 0.1$, and $\sigma_Z = 0.01$. The optimization method was repeated 25 times with random re-initialization for each run to obtain:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}_\theta(v_{1:n}). \quad (5.15)$$

and we proceeded in the same way with 50 different trajectories to obtain the following results. See Tables 5.1, and 5.2: this shows that the algorithm is able to obtain quite good estimates of the parameters for simulated data.

5.5 Parameter estimation based on historical data

For our empirical analysis we again use the data from the European Energy Exchange (EEX) in Leipzig, Germany. In our analysis we considered the spot price of the EEX baseload index and monthly baseload futures prices. The spot price is an equally weighted average of all 24 hourly spot prices for that particular day. Holidays and weekends have been removed from the data set.

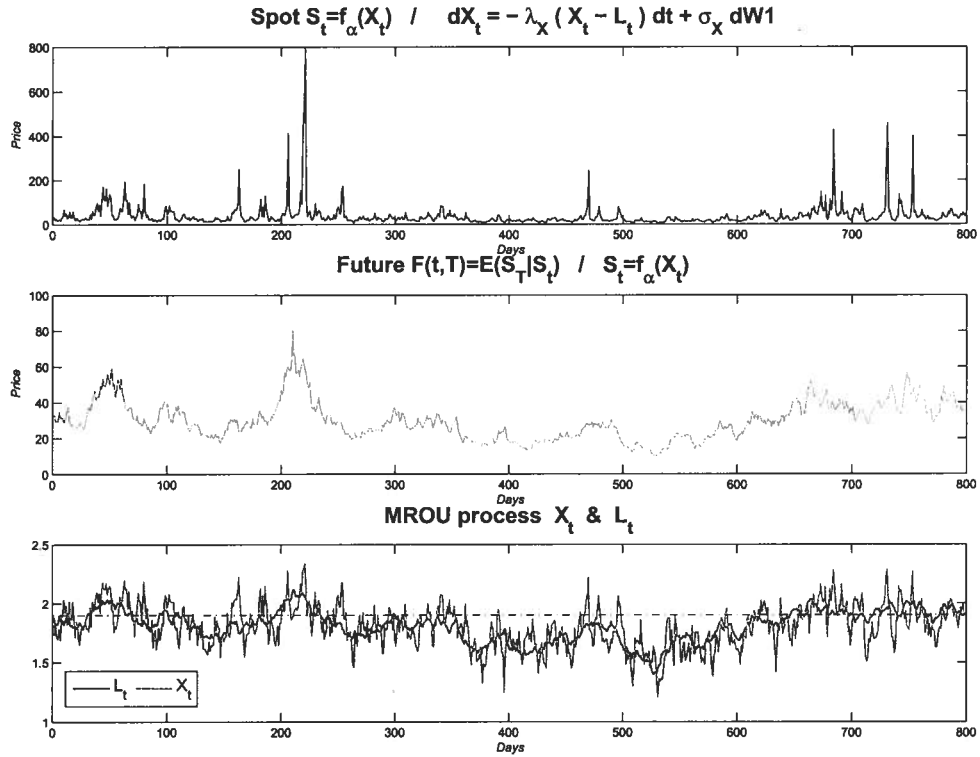


Figure 5.1: The graph shows a simulated trajectory of the NLMROU model.

$n = 800$						
	True Value	Run 1	Run 2	Run 3	Run 4	Run 5
λ_X	100	102.562	105.627	101.162	104.041	106.12
λ_L	3	3.523	2.871	3.466	3.540	3.618
σ_X	1.7	1.582	1.530	1.845	1.728	1.652
σ_L	0.4	0.419	0.457	0.420	0.380	0.379
L	1.9	1.951	1.875	1.952	1.803	1.974
α	-0.4	-0.402	-0.382	-0.461	-0.389	-0.406
$-\mathcal{L}_t(\theta)$		-186.276	-179.261	-197.417	-186.248	-191.87

Table 5.1: Five different maximization runs with 800 observations.

Our data comprise almost five years baseload day prices from July 1, 2002 to June 29, 2007, totaling 1267 observations. The dynamics of the spot prices for the considered period are shown in Figure 3.3.

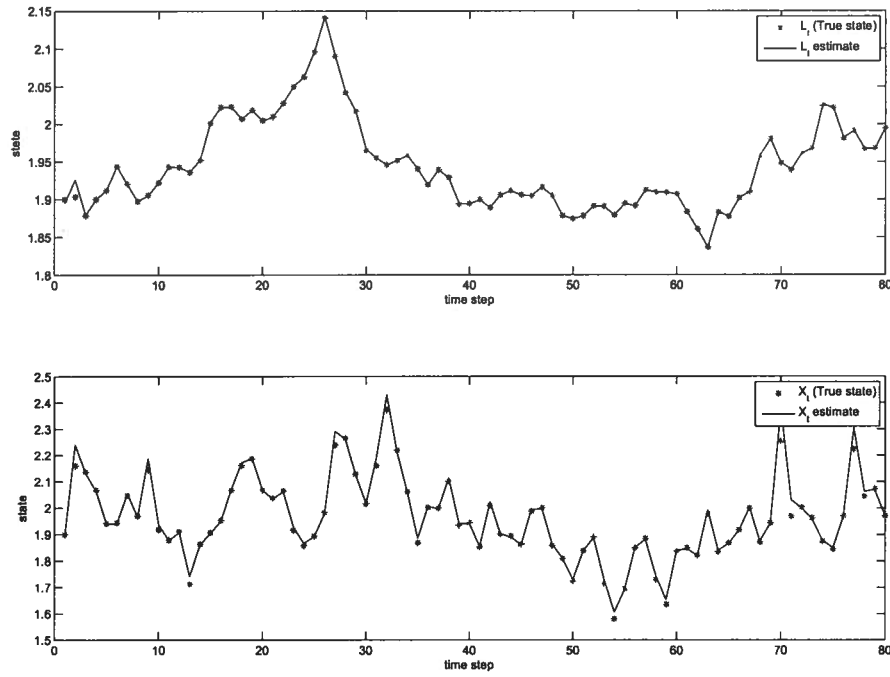


Figure 5.2: Plot of the true and estimated processes L_t and X_t of the NLMROU model (first 80 observations).

	$n = 800$		
	True value	Estimator	Std.
λ_X	100	102.4069	4.340
λ_L	3	2.964	0.795
σ_X	1.7	1.786	0.216
σ_L	0.4	0.3954	0.0921
L	1.9	1.8968	0.0216
α	-0.4	-0.4009	0.038

Table 5.2: Estimation using one futures contract ($n = 800$).

To estimate the parameters of the deterministic part

$$h(t) = \eta + \beta_0 t + \beta_1 \cos\left(\frac{\tau_1 + 2\pi t}{250}\right) + \beta_2 \cos\left(\frac{\tau_2 + 2\pi t}{5}\right),$$

we ran the least squares method as we did in Chapter 3. See Table 3.5 for the estimated values of η , β_0 , β_1 , β_2 , τ_1 , and τ_2 .

In the following we will analyze the three time series of the EEX market for the periods July 1, 2002 - December 31, 2004, January 1, 2005 - June 29, 2007, and the whole series. The seasonalities have been removed.

For simplicity and to reduce the number of parameters to estimate, we fixed the noise parameters to be constants. $\sigma_S = 0.1$ and $\sigma_Z = 0.01$. In view of the low correlation found between the processes X_t and L_t in Chapter 3, we took the covariance $\rho = 0$.

The results on the parameter estimates are shown in the Table 5.3.

	Part 1	Part 2	Whole
	Estimator	Estimator	Estimator
λ_X	127.599	116.036	121.557
λ_L	0.11178	0.753711	0.203681
σ_X	1.25847	0.756485	1.15815
σ_L	0.0834349	0.0730882	0.0901108
L	1.86049	1.77443	1.79335
α	-0.438183	-0.489661	-0.414042

Table 5.3: Estimated values for the EEX market using S_t , and $F(t, T_1)$.

The parameters λ_X and λ_L relate, as in Chapter 3, to the ‘half-life’ of the mean reversion of the short-term and long-term processes. We obtained similar results for $\tau_X = \log 2 / \lambda_X$, showing a half life of 1 – 2 days. The estimated half-lives for L_t differ significantly for the two periods, being about 6 years for the first period, and about 8 months for the second. While the MROU and NLMROU models give similar estimates for the speed of mean reversion of the short-term component X_t , the NLMROU model gives significantly slower mean reversion for the long-run process L_t .

The estimates of the nonlinearity parameter α are quite similar for the two periods.

As in the case of the MROU process, comparison of the variances $V_L = \sigma_L^2 / 2\lambda_L$, $V_X = \sigma_X^2 / 2\lambda_X$ show that the dynamics of X_t and L_t contribute significantly to the long run variation of S_t . As before the long run process

gives a somewhat greater contribution. We obtain the following results:

	$\sqrt{V_L}$	$\sqrt{V_X}$
Period 1	0.18	0.08
Period 2	0.06	0.05

Note that because we are applying a different function ($f_\alpha(\cdot)$ rather than $\exp\{\cdot\}$) to X_t to obtain the spot price, it does not make sense to directly compare the values of σ_X and σ_L with those obtained in Chapter 3.

In order to investigate if the estimated parameters make sense, we simulated a path of the spot price, future prices and long and short term mean process that describe the model using the estimated values from Table 5.3 considering the whole data, see Figure 5.3. Empirical moments of the EEX spot price versus simulated moments (averaged over 50 simulation paths) are shown in Table 5.4.

Comparing the real spot price with the price produced by the model, we see that the model is able to produce significant price spikes with values similar to those for the real data. As with the real data, the spikes tend to bunch together. As well, the model exhibits periods of high variance - compare for example the period 700 – 950 in Figure 3.5, and 750 – 950 in Figure 5.3. The estimated parameters for the long-run mean process give quite slow mean reversion, and Figure 5.3 shows a process of this kind.

This model will tend to generate more spikes when the L_t process is large, and one sees this at the end of the simulation, when the simulated spot has many small spikes in the range from €150 on. Similar features are seen in the real data - in the period 800 to 1100 in Figure 3.5.

The model does however have some defects. The first is that the skewness of the log returns is close to zero - see Table 5.4 (this is likely to be a feature of any model without jumps). Then this model is not able to capture the skewness which is present in the real data. It also appears that the model underestimates the kurtosis, compared with the real data.

In spite of these problems, this model appears to offer a significant improvement over the MROU model at a quite moderate ‘cost’ - the cost being in terms of additional parameters and complications of parameter estimation. Although the model is far from perfect, its performance suggests that it is

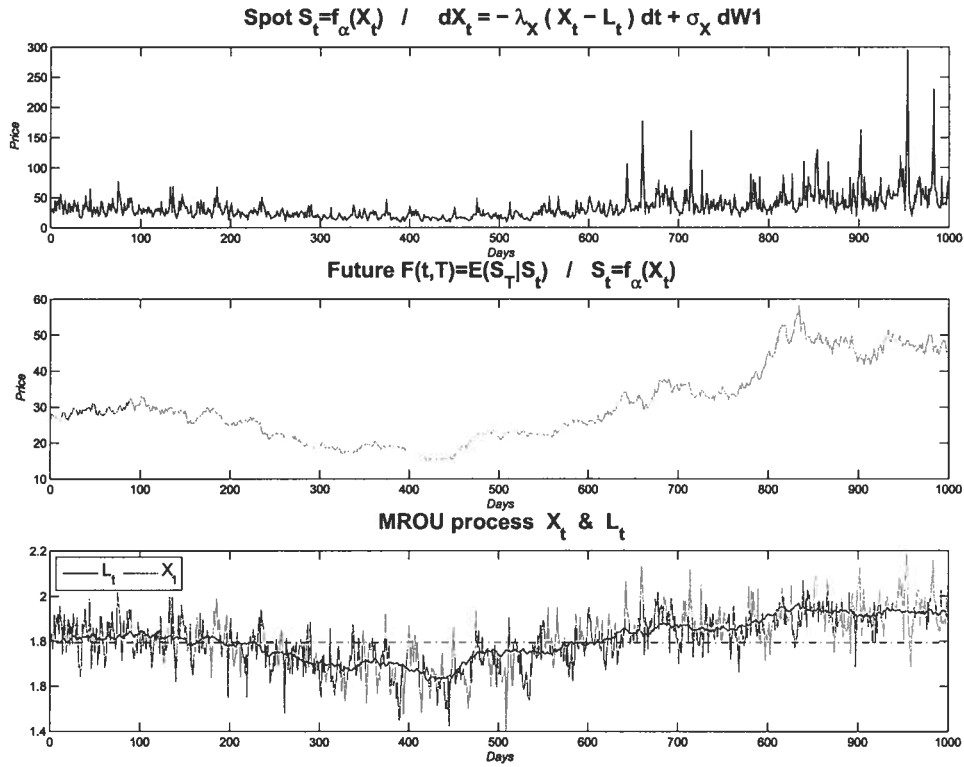


Figure 5.3: Simulation of spot and future prices price using estimated values for the whole data set.

	Real data (1)	Sim. data (1)	Real data (2)	Sim. data (2)
Mean	0.0019	0.0000	0.0000	-0.00037
Std Dev.	0.2214	0.60029	0.2066	0.46088
Skewness	-1.3125	-0.02066	-0.0800	0.04927
Kurtosis	29.8250	4.77539	8.1288	3.99024
n data	634	634	631	631

Table 5.4: The table shows the first four moments of the logarithmic de-seasonalized price returns of observed data and the average of 50 simulated trajectories.

well worth considering more elaborate models of this type in the search for a good description of electricity prices.

Chapter 6

Conclusions

In this thesis, based on the specific properties of electricity we have proposed three process models that incorporate various features of power prices. We calibrate these models, using both spot and futures prices, to artificial and historical data applying three filtering methods.

The first model is a two-factor linear Gaussian model. This models the log-spot price as mean-reverting process, where the mean reversion is to a second (unobservable) “long run mean” process. This second process is modeled by an Ornstein-Uhlenbeck process. In this thesis (unlike the work reported in [4]) we used both the spot and future prices to estimate parameters. The space-state formulation of this model is suitable to the application of Kalman filtering techniques, and we used a maximum likelihood estimator based on the Kalman filter. This worked well for simulated data, and we then applied it to estimate parameters for the German EEX market. Simulations suggest that this model, with the fitted parameters, does fit some features of the real data. However, it definitely fails to exhibit some features of the real data, such as the jumps or price spikes seen in Figure 3.3.

This defect in the first model led us to consider a second model, which incorporates jumps. We kept the same basic form of a log-spot price, and a long-term mean process, but added jumps to the log-spot price. For simplicity we took the distribution of both the upward and downward jumps to be exponential, with possibly different rates and parameters. The standard Kalman filter cannot be applied satisfactory in this case, since the model is non-Gaussian. One alternative, which requires rather weak assumptions on the distribution involved, is the particle filter. We developed code to use the particle filter for the second (jump diffusion) model based on the kernel approximation of the posteriori suggested by Liu & West (2001). An empirical application on simulated data is presented to study the performance of the implemented algorithm. In general we observed that while the particle filter can work satisfactorily in estimating the unknown parameters, it is very sen-

sitive to the particular form of dynamics for the artificial parameters used in the parameter estimation. In view of this it is hard to apply to real data.

The third model we presented is an extension of the nonlinear Ornstein-Uhlenbeck model (NLOU) proposed in Barlow [3]. While Barlow used only one factor to describe the dynamics of the spot price in this thesis we consider a two-factor model and the same nonlinear transformations to model the spot price. The model captures the mean-reversion, jumps and spikes behavior observed in real market. The model has the advantage over jump-diffusion models that it is Gaussian. Hence we can use the unscented Kalman filter algorithm to estimate the NLMROU model. We calibrate the models to daily EEX market obtaining similar simulated trajectories with the estimated parameters.

6.1 Future work

1. In Chapter 5 we analyzed a NLMROU model of the form

$$\tilde{S}_t = h(t)f_\alpha(X_t)$$

where (X_t, L_t) is given by (5.2)-(5.3). The original model in [3] was justified by considerations of supply and demand curves, and so it might be more realistic to consider a model of the form

$$\tilde{S}_t = f_\alpha(X_t + h(t)) \tag{6.1}$$

where $h(t)$ is a deterministic seasonal correction. This model would have the merit of generating spikes during periods of high demand, without the need to consider different regimes, as is done in [45, 72].

The main obstacle to parameter estimation for (6.1) is that, since α is unknown, it is no longer possible to estimate $h(t)$ by a least squares method.

One possible approach is an iterative method. One would first apply the UKF to the uncorrected series $(\tilde{S}_t, \tilde{F}(t, T))$, to obtain an initial estimate $\hat{\alpha}_1$. One then applies least squares to the series $f_{\hat{\alpha}_1}^{-1}(\tilde{S}_t)$, to obtain

an estimate \hat{h}_1 . Given \hat{h}_1 , the spot component of the measurement equation is

$$v^{(1)}(t) = \log \left(f_{\alpha}(X_t + \hat{h}(t)) \right),$$

which can be calculated as in Section 5.2. Using this, and a similar expression for the futures component, one can then apply the UKF to obtain a second round of parameter estimates, and in particular an improved estimate $\hat{\alpha}_2$ for α . Iterating, one would hope to obtain estimates for α and $h(t)$.

This procedure seems feasible to implement, but its convergence properties and stability are at this point not clear.

2. Improved parameter estimation using the particle filter.

In this thesis we applied the Bayesian approach where an augmented state variable which includes the parameters is processed by the particle filter for the NLMROU model. We adopted the on-line estimation of parameters and state developed by Liu & West (2001). The main feature of this approach is that the variance matrix Λ_t shrinks step by step and it finally converges towards 0. Hence the parameters could converge towards a wrong value because the Λ_t converges towards 0 before reaching the true parameter value. This problem cannot be avoided when we do not have prior knowledge about the parameters. That is why the method is very sensitive to the initial values of the added noise in the parameters.

Along the same lines, another method that can we use is the ‘practical filter’ proposed by Polson et al. (2008) [46, 67]. Their approach is based on approximating the target posterior by a mixture of fixed-lag smoothing distributions. According to the authors, unlike the particle filter approaches, it provides independent samples from the target distribution, does not suffer from particle degeneracies, and handles outliers and high dimensional problems well.

Meyer-Brandis T. & Tankov P. (2008) [62] comment:

“Sequential filtering makes less sense when the complete series is available for estimation”.

Although computationally more intensive, the difficulties encountered in implementing the particle filter suggest this may be the correct approach. An approach used by Olsson et al. (2008) [65] and Wills et al. (2008) [81] is an off-line method performing maximum likelihood estimation via the EM algorithm. An essential component in the E-step is to approximate the ‘smoothing distribution’, that is $\{p_\theta(u_t|v_{1:n}); t = 1, \dots, n\}$. In the general nonlinear and non-Gaussian case various schemes have been proposed. The fixed-lag approximation is the simplest approach and it was first proposed in [54]. In [65] they apply the particle filter technique to smooth additive functionals based on a fixed-lag smoother. The method exploits the forgetting properties on the conditional hidden chain and is not affected by the degeneracy of the particle trajectories.

Approaches of this kind are well worth investigating for jump models of the type considered in Chapter 4.

Bibliography

- [1] L. Anderson and D. Davison. A hybrid system-econometric model for electricity spot prices: Considering spike sensitivity to forced outage distributions. *IEEE Transactions on Power Systems*, 23(3):927–937, 2008.
- [2] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEE Trans. Signal Processing*, 50(2):174–188, 2002.
- [3] M. Barlow. A diffusion model for electricity prices. *Mathematical Finance*, 12:287–298, 2002.
- [4] M. Barlow, M. Gusev, and M. Lai. Calibration of multifactor models in electricity markets. *International Journal of Theoretical and Applied Finance*, 7(2):101–120, 2004.
- [5] D.R. Beaglehole and M.S. Tenney. General solutions of some interest rate contingent claim pricing equations. *Journal of Fixed Income*, 1:69–83, 1991.
- [6] F.E. Benth, J.S. Benth, and S. Koekebakker. *Stochastic Modelling of Electricity and Related Markets*. World Scientific, 2008.
- [7] F.E. Benth, J. Kallsen, and T. Meyer-Brandis. A non-Gaussian Ornstein-Uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, 14(2):153–169, 2007.
- [8] F.E. Benth and S. Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.
- [9] T. Bjork. *Arbitrage Theory in Continuous Time, 2nd Ed.* Oxford University Press, 2004.

Bibliography

- [10] S. Borovkova and H. Geman. Analysis and modelling of electricity futures prices. *Studies in Nonlinear Dynamics & Econometrics*, 10(3):Article 6, 2006.
- [11] M.J. Brennan and E. Schwartz. Evaluating natural resource investments. *Journal of Business*, 1985.
- [12] M. Burger, B. Klar, A. Mueller, and G. Schindlmayr. A spot market model for pricing derivatives in electricity markets. *Journal of Quantitative Finance*, 4:109–122, 2004.
- [13] O. Cappe, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer, 2006.
- [14] R. Carmona and M. Ludkovski. Spot convenience yield models for the energy markets. *AMS Mathematics of Finance*, 351:65–80, 2004.
- [15] C. Carraro. Square root Kalman algorithms in econometrics. *Computer Sciences in Economics and Management*, 1:41–51, 1988.
- [16] A. Cartea and G. Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.
- [17] A. Cartea and P. Villaplana. Spot price modeling and the valuation of electricity forward contracts. *Journal of Banking and Finance*, 32(12):2502–2519, 2008.
- [18] T. Chen, J. Morris, and E. Martin. Particle filters for state and parameter estimation in batch process. *Journal of Process Control*, pages 665–673, 2005.
- [19] L. Clewlow and C. Strickland. Valuing energy options in a one factor model fitted to forward prices. *Working Paper, University of Sydney*.
- [20] L. Clewlow and C. Strickland. *Energy Derivatives: Pricing and Risk Management*. Lacima Group, 2000.
- [21] M. Culot, V. Goffin, S. Lawford, S. de Menten, and Y. Smeers. An affine jump diffusion model for electricity. *working paper*, 2006.

Bibliography

- [22] D. Davison, L. Anderson, B. Marcus, and K. Anderson. Development of a hybrid model for electrical power spot prices. *IEEE Transactions on Power Systems*, 17(2):257–264, 2002.
- [23] J.F.G. De Freitas, M. Niranjana, and A.H. Gee. Dynamic learning with the EM algorithm for neural networks. *Journal of VLSI Signal Processing Systems*, 26:119 – 131, 2000.
- [24] S. Deng. Pricing electricity derivatives under alternative stochastic spot price models. *33rd Hawaii International Conference on System Sciences*, 4:4025–4034, 2000.
- [25] J.E. Jr. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, 1996.
- [26] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.
- [27] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 6:1343–1376, 2000.
- [28] J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.
- [29] C. Erlwein, F. Benth, and R.S. Mamon. Hmm filtering and parameter estimation of electricity spot price model. *Revised for Energy Economics*, 2009.
- [30] A. Eydeland and K. Wolyniec. *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging*. John Wiley & Sons, 2003.
- [31] H. Geman. *Commodities and Commodity Derivatives*. John Wiley & Sons, 2005.
- [32] H. Geman and A. Roncoroni. Understanding the fine structure of electricity prices. *Journal of Business*, 79(2):1225–1261, 2006.
- [33] H. Geman and O. Vasicek. Forwards and futures on non storable commodities. *RISK*, August:93–97, 2001.

Bibliography

- [34] Z. Ghahramani and S. Roweis. Learning nonlinear dynamical systems using an EM algorithm. *Advances in Neural Information Processing Systems*, 11:599–605, 1999.
- [35] R. Gibson and E. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 3:959–976, 1990.
- [36] B. Gopaluni, T.B. Schon, and A. Wills. Particle filter approach to non-linear system identification under missing observations with a real application. *Proceedings of IFAC Symposium on System Identification*, 2009.
- [37] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEE Proceeding -F Radar, Sonar Navig.*, 140(2):107–113, 1993.
- [38] M. Grewal and A.P. Andrews. *Kalman filtering: Theory and practice using MATLAB*. Wiley, 2001.
- [39] J.M. Griffin and S.L. Puller. *Electricity deregulation: choices and challenges*. Series in the Economics of Public Policy, 2005.
- [40] J. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [41] S. Haykin. *Kalman Filtering and Neural Networks*. John Wiley & Sons, 2001.
- [42] D. Heath, R. Jarrow, and A. Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claim valuation. *Econometrica*, 60:77–105, 1992.
- [43] T. Higuchi. Self-organizing time series model. *Sequential Monte Carlo Methods in Practise (Doucet, de Freitas and Gordon, Eds)*, pages 429–444, 2001.
- [44] S. Hikspoors and S. Jaimungal. Energy spot price models and spread options pricing. *International Journal of Theoretical and Applied Finance*, 10(7):1111–1135, 2007.
- [45] R. Huisman and R. Mahieu. Regime jumps in electricity prices. *Energy Economics*, 25:425–434, 2003.

Bibliography

- [46] M.S. Johannes, N.G. Polson, and J.R. Stroud. Optimal filtering of jump diffusions: Extracting latent states from asset prices. *Review of Financial Studies*, 22(7):2559–2599, 2009.
- [47] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–421, 2004.
- [48] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, D 82:35–45, 1960.
- [49] T. Kanamura and K. Ohashi. A structural model for electricity prices with spikes : Measurement of spike risk and optimal policies for hydropower plant operation. *Energy economics*, 29(5):1010–1032, 2007.
- [50] N. Kantas, A. Doucet, S.S. Singh, and J.M. Maciejowski. Overview of Sequential Monte Carlo methods for parameter estimation on general state space models. *15th IFAC Symposium on System Identification (SYSID)*, In Press, 2009.
- [51] B.P. Kellerhals. *Financial Pricing Models in Continuous Time and Kalman Filtering*. Springer, 2001.
- [52] V.A. Kholodnyi. A non-markovian process for power prices with spikes and valuation of european contingent claims of power. *Preprint*, 2001.
- [53] G. Kitagawa. Self-organizing state space model. *Journal of the American Statistical Association*, 93(443):1203–1215, 1998.
- [54] G. Kitagawa and S. Sato. Monte carlo smoothing and self-organizing state-space models. *Sequential Monte Carlo Methods in Practise (Doucet, de Freitas and Gordon, Eds)*, 2001.
- [55] S. Koekebakker and F. Ollmar. Forward curve dynamics in the Nordic electricity market. *Managerial Finance*, 31(6):73–94, 2005.
- [56] A. Leon and A. Rubia. Testing for weekly seasonal unit roots in the Spanish power pool. *Modelling prices in competitive electricity markets, D.W. Bunn Ed*, pages 177–189, 2004.

Bibliography

- [57] J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. *Sequential Monte Carlo Methods in Practise* (Doucet, de Freitas and Gordon, Eds), pages 197–223, 2001.
- [58] J.S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.
- [59] J. Lucia and E. Schwartz. Electricity prices and power derivatives - evidence from the Nordic power exchange. *Review of Derivatives Research*, 5:5–50, 2002.
- [60] M.R. Lyle and R.J. Elliott. A simple hybrid model for power derivatives. *Energy Economics*, 31, 2009.
- [61] M. Manoliu and S. Tompaidis. Energy futures prices: term structure models with Kalman filter estimation. *Apply Mathematical Fiance*, 9:21–43, 2002.
- [62] T. Meyer-Brandis and P. Tankov. Multi-factor jump-diffusion models of electricity prices. *International Journal of Theoretical and Applied Finance*, 11(5):503–528, 2008.
- [63] A. Misiorek, S. Trueck, and R. Weron. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Studies in Nonlinear Dynamics and Econometrics*, 10, 2006.
- [64] N.K. Nomikos and O. Soldatos. Using affine jump diffusion models for modelling and pricing electricity derivatives. *Applied Mathematical Finance*, 15(1):4171, 2008.
- [65] J. Olsson, O. Cappé, R. Douc, and E. Moulines. Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli*, 14, 2008.
- [66] D. Pilipovic. *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill, 1998.
- [67] N.G. Polson, J. R. Stroud, and P. Müller. Practical filtering with sequential parameter learning. *Journal of the Royal Statistical Society, Series B*, 70(2):413–428, 2008.

Bibliography

- [68] D. Raggi. Adaptive MCMC methods for inference on affine stochastic volatility models with jumps. *The Econometric Journal*, 8(2):235–250, 2005.
- [69] B. Ristic, S. Aralampalam, and N. Gordon. *Beyond the Kalman filter*. Artech House, 2004.
- [70] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods, 2nd Ed.* Springer-Verlag, 2004.
- [71] S. Roweis and Z. Ghahramani. A unifying review of linear Gaussian models. *Neural Computation*, 11:305–345, 1999.
- [72] G. Schindlmayr. A regime-switching model for electricity spot prices. *Working Paper*, 2005.
- [73] E.S. Schwartz. The stochastic behavior of commodity prices: implications for valuation and hedging. *Journal of Finance*, 52(3):923–973, 1997.
- [74] R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications: With R Examples, 2nd Ed.* Springer, 2006.
- [75] D. Simon. *Optimal State Estimation*. John Wiley & Sons, 2006.
- [76] A.F.M. Smith and A.E. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2):84–88, 1992.
- [77] H. Tanizaki. *Nonlinear Filters: Estimation and Applications*. Springer-Verlag, 1996.
- [78] J. Teichmann. A note on nonaffine solution of term structure equations with applications to power exchanges. *Mathematical Finance*, 15(1):191–201, 2005.
- [79] A. van der Merwe, R. Doucet and N. de Freitas. The unscented particle filter. *Technical Report CUED*, 380, 2000.
- [80] P. Villaplana. Pricing power derivatives: a two-factor jump-diffusion approach. *Business Economics Series (Working Paper)*, 2003.

Bibliography

- [81] A. Wills, T. Schon, and B. Ninness. Parameter estimation for discrete-time nonlinear systems using EM. *Proceedings of the 17th IFAC World Congress, Seoul, Korea*, 2008.
- [82] L. Xiong. *Stochastic models for electricity prices in Alberta*. University of Calgary, M.Sc. thesis, 2004.

Appendix A

An Affine Jump-Diffusion process (AJD) is a jump-diffusion process for which the drifts and covariances and jump intensities are linear in the state vector \mathbf{U} . Duffie et al. (2000) [27] show that AJD processes are analytically tractable in general.

Let \mathbf{U} be a strong Markov process with realizations $\{\mathbf{U}_t, 0 \leq t < \infty\}$ in some state space $D \subset \mathbb{R}^n$, which solves the following stochastic differential equation

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

The jump behavior of \mathbf{U} is governed by m types of jump processes. Each jump type Z_t^i is a pure jump process with a stochastic arrival intensity $\lambda^i(\mathbf{U}_t, t)$ for some $\lambda^i : (D, t) \mapsto \mathbb{R}^n$ and jump amplitude distribution v_t^i on \mathbb{R}^n , where v_t^i only depends on time t . The functions $\mu : (D, t) \mapsto \mathbb{R}^n$ and $\sigma : (D, t) \mapsto \mathbb{R}^{n \times n}$ are assumed to be Lipschitz continuous in order to guarantee that (A-1) has a unique solution. The process \mathbf{W}_s is a standard Brownian motion in \mathbb{R}^n .

The process \mathbf{U}_t defined by (A-1), is said to be an affine jump-diffusion process if

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

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

Appendix A

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. Finally $l_0^i(t) \in \mathbb{R}$.

Notice that given an initial condition \mathbf{U}_0 , the tuple $(\mathbf{K}_0, \mathbf{K}_1, \mathbf{H}_0, \mathbf{H}_1, l_0)$ can be used to determine a transform $\Psi : \mathbb{C}^n \times [0, \infty) \times [0, \infty) \times D \mapsto C$ of \mathbf{U}_T conditional on \mathbf{U}_t , $0 \leq t \leq T$, defined by

$$\Psi(\mathbf{u}, t, T, \mathbf{U}_t) := \mathbb{E}[\exp\{\mathbf{u} \cdot \mathbf{U}_T\} | \mathbf{U}_t] \quad (\text{A-2})$$

where \mathbb{E} denotes the expectation under the distribution of \mathbf{U}_T determined by $(\mathbf{K}_0, \mathbf{K}_1, \mathbf{H}_0, \mathbf{H}_1, l_0)$. If we suppose $(\mathbf{K}_0, \mathbf{K}_1, \mathbf{H}_0, \mathbf{H}_1, l_0)$ is well-behaved at (\mathbf{u}, T) , then the transform Ψ of \mathbf{U}_t , $0 \leq t \leq T$, defined by (A-2) exists and is given by:

$$\Psi(\mathbf{u}, t, T, \mathbf{U}_t) = \exp\{M(\mathbf{u}, t, T) + \mathbf{N}(\mathbf{u}, t, T) \cdot \mathbf{U}_t\}. \quad (\text{A-3})$$

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

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

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

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

Here $\varphi_i(\mathbf{c})$ is the “jump transform” for the i -th jump. It is given by

$$\varphi_i(\mathbf{c}) = \int_{\mathbb{R}^n} \exp\{\mathbf{c} \cdot \mathbf{U}\} d\nu_t^i(\mathbf{U}) \quad (\text{A-4})$$

whenever the integral is well defined.

Appendix A

We define the extended transform $\Phi : \mathbb{R}^n \times \mathbb{C}^n \times [0, \infty) \times [0, \infty) \times D \mapsto \mathbb{C}$ of \mathbf{U}_T conditional on \mathbf{U}_t by

$$\Phi(\mathbf{v}, \mathbf{u}, t, T, \mathbf{U}_t) := E[(\mathbf{v} \cdot \mathbf{U}_T) \exp\{\mathbf{u} \cdot \mathbf{U}_T\} | \mathbf{U}_t]. \quad (\text{A-5})$$

Given sufficient regularity the “extended transform” Φ can be computed by differentiation of the transform Ψ . Hence

$$\Phi(\mathbf{v}, \mathbf{u}, t, T, \mathbf{U}_t) = \Psi(\mathbf{u}, t, T, \mathbf{U}_t) \{ \mathbf{C}(t) + \mathbf{D}(t) \cdot \mathbf{U}_t \} \quad (\text{A-6})$$

where Ψ is given by (A-3), and $\mathbf{C}(\cdot)$ and $\mathbf{D}(\cdot)$ satisfy

$$\frac{\partial \mathbf{C}(\cdot)}{\partial t} = -\mathbf{K}_0(t)' \mathbf{D} - \mathbf{N}' \mathbf{H}_0(t) \mathbf{D} - l_0(t) \nabla \varphi(\mathbf{N}) \mathbf{D}, \quad (\text{A-7})$$

$$\frac{\partial \mathbf{D}(\cdot)}{\partial t} = -\mathbf{K}_1(t)' \mathbf{D} - \mathbf{N}' \mathbf{H}_1(t) \mathbf{D}, \quad (\text{A-8})$$

with the boundary conditions

$$\mathbf{C}(T) = 0, \quad \mathbf{D}(T) = \mathbf{v}. \quad (\text{A-9})$$

Here $\nabla \varphi(\mathbf{c})$ is the gradient of $\varphi(\mathbf{c})$ with respect to $\mathbf{c} \in \mathbb{C}^n$.

Appendix B

In this appendix we solve the ODEs (3.13)-(3.15) which arise in the calculation of futures prices in the AJD model.

We begin with the ODEs arising from the MROU model. Recall that

$$\frac{\partial N_1}{\partial t} = \lambda_X N_1, \quad (\text{B-1})$$

$$\frac{\partial N_2}{\partial t} = -\lambda_X N_1 + \lambda_L N_2, \quad (\text{B-2})$$

$$\frac{\partial M}{\partial t} = -\lambda_L L N_2 - \frac{1}{2}(\sigma_X^2 N_1^2 + \sigma_L^2 N_2^2) - N_1 N_2 \rho \sigma_X \sigma_L. \quad (\text{B-3})$$

We begin with (B-1). We have $N_1 = N_1(t, T)$ and

$$\frac{\partial N_1}{\partial t} = \lambda_X N_1, \quad N_1(u_1, T, T) = u_1. \quad (\text{B-4})$$

If we fix T , so regard N_1, N_2 as function of t only, then (B-1)-(B-3) are ODEs and (B-4) has solution

$$N_1 = e^{\lambda_X t} A.$$

Since

$$u_1 = e^{\lambda_X T} A, \quad A = u_1 e^{-\lambda_X T},$$

therefore

$$N_1(u_1, t, T) = u_1 e^{\lambda_X(t-T)}. \quad (\text{B-5})$$

We now treat (B-3):

$$\frac{dN_2}{dt} = -\lambda_X N_1 + \lambda_L N_2, \quad N_2(u_2, T, T) = u_2. \quad (\text{B-6})$$

Appendix B

Substituting for N_1 , and multiplying by a factor μ we obtain

$$\mu \frac{dN_2}{dt} - \mu \lambda_L N_2 = -\lambda_X u_1 e^{\lambda_X(t-T)} \mu. \quad (\text{B-7})$$

So

$$\frac{(\mu N_2)}{dt} = \mu \frac{dN_2}{dt} - \mu \lambda_L N_2,$$

$$\frac{d\mu}{dt} N_2 = -\mu \lambda_L N_2,$$

$$\frac{d\mu}{dt} = -\mu \lambda_L,$$

$$\frac{d\mu}{\mu} = -\lambda_L dt,$$

$$\mu = e^{-\lambda_L t}.$$

Substituting for μ in equation (B-7), we obtain

$$\frac{d(e^{-\lambda_L t} N_2)}{dt} = -\lambda_X u_1 e^{\lambda_X(t-T)} e^{-\lambda_L t} \quad (\text{B-8})$$

$$= -\lambda_X u_1 e^{t(\lambda_X - \lambda_L)} e^{-\lambda_X T} \quad (\text{B-9})$$

so,

$$e^{-\lambda_L t} N_2 = -\frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{t(\lambda_X - \lambda_L)} e^{-\lambda_X T} + C, \quad (\text{B-10})$$

$$N_2 = -\frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{\lambda_X(t-T)} + C e^{\lambda_L t}. \quad (\text{B-11})$$

Now, solving for C

$$u_2 = N_2(u_2, T, T) = -\frac{\lambda_X u_1}{\lambda_X - \lambda_L} + C e^{\lambda_L T} \quad (\text{B-12})$$

therefore

$$C = u_2 e^{-\lambda_L T} + \frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{-\lambda_L T}. \quad (\text{B-13})$$

Substituting C in equation (B-11) gives

$$N_2(u_2, t, T) = -\frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{\lambda_X(t-T)} + e^{\lambda_L t} (u_2 e^{-\lambda_L T} + \frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{-\lambda_L T}) \quad (\text{B-14})$$

$$= -\frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{\lambda_X(t-T)} + u_2 e^{\lambda_L(t-T)} + \frac{\lambda_X u_1}{\lambda_X - \lambda_L} e^{\lambda_L(t-T)}. \quad (\text{B-15})$$

Finally we solve (B-3).

$$\frac{dM}{dt} = -\lambda_L L N_2 - \frac{1}{2}(\sigma_X^2 N_1^2 + \sigma_L^2 N_2^2) - N_1 N_2 \rho \sigma_X \sigma_L \quad (\text{B-16})$$

with $M((u_1, u_2), T, T) = 0$.

Let $\alpha = -\lambda_X u_1 / (\lambda_X - \lambda_L)$. Replacing the solution for N_1 and N_2 given by equations (B-5) and (B-15) in (B-16) we have

$$\begin{aligned} \frac{dM}{dt} &= -\lambda_L L (\alpha e^{\lambda_X(t-T)} + u_2 e^{\lambda_L(t-T)} - \alpha e^{\lambda_L(t-T)}) - \frac{1}{2} [\sigma_X^2 u_1^2 e^{2\lambda_X(t-T)} \\ &\quad + \sigma_L^2 (\alpha e^{\lambda_X(t-T)} + u_2 e^{\lambda_L(t-T)} - \alpha e^{\lambda_L(t-T)})^2] \\ &= -\lambda_L L \alpha e^{\lambda_X(t-T)} - \lambda_L L u_2 e^{\lambda_L(t-T)} + \lambda_L L \alpha e^{\lambda_L(t-T)} \\ &\quad - \frac{\sigma_X^2 u_1^2}{2} e^{2\lambda_X(t-T)} - \frac{\sigma_L^2 \alpha^2}{2} e^{2\lambda_X(t-T)} - \frac{\sigma_L^2 u_2}{2} e^{2\lambda_L(t-T)} \\ &\quad - \frac{\sigma_L^2 \alpha^2}{2} e^{2\lambda_L(t-T)} - \sigma_L^2 \alpha u_2 e^{\lambda_X(t-T)} + \sigma_L^2 u_2 \alpha e^{2\lambda_L(t-T)} \\ &\quad + \sigma_L^2 \alpha^2 e^{(\lambda_X + \lambda_L)(t-T)} - \rho \sigma_X \sigma_L \alpha u_1 e^{\lambda_X(t-T)} \\ &\quad + \rho \sigma_X \sigma_L u_1 u_2 e^{(\lambda_X + \lambda_L)(t-T)} + \rho \sigma_X \sigma_L \alpha u_1 e^{(\lambda_X + \lambda_L)(t-T)}. \end{aligned}$$

Appendix B

Integrating both sides gives

$$\begin{aligned}
M = & -\frac{\lambda_L L \alpha}{\lambda_X} e^{\lambda_X(t-T)} - \frac{\lambda_L L u_2}{\lambda_L} e^{\lambda_L(t-T)} + \frac{\lambda_L L \alpha}{\lambda_L} e^{\lambda_L(t-T)} \\
& - \frac{\sigma_X^2 u_1^2}{4\lambda_X} e^{2\lambda_X(t-T)} - \frac{\sigma_L^2 \alpha^2}{4\lambda_X} e^{2\lambda_X(t-T)} - \frac{\sigma_L^2 u_2}{4\lambda_L} e^{2\lambda_L(t-T)} \\
& - \frac{\sigma_L^2 \alpha^2}{4\lambda_L} e^{2\lambda_L(t-T)} - \frac{\sigma_L^2 \alpha u_2}{\lambda_X} e^{\lambda_X(t-T)} + \frac{\sigma_L^2 u_2 \alpha}{2\lambda_L} e^{2\lambda_L(t-T)} \\
& + \frac{\sigma_L^2 \alpha^2}{\lambda_X + \lambda_L} e^{(\lambda_X + \lambda_L)(t-T)} - \frac{\rho \sigma_X \sigma_L \alpha u_1}{2\lambda_X} e^{\lambda_X(t-T)} \\
& - \frac{\rho \sigma_X \sigma_L \alpha u_1 u_2}{\lambda_X + \lambda_L} e^{(\lambda_X + \lambda_L)(t-T)} + \frac{\rho \sigma_X \sigma_L \alpha u_1}{\lambda_X + \lambda_L} e^{(\lambda_X + \lambda_L)(t-T)} + C.
\end{aligned}$$

Since $M((u_1, u_2), t, T) = 0$,

$$\begin{aligned}
M = & -\frac{\lambda_L L \alpha}{\lambda_X} (e^{\lambda_X(t-T)} - 1) - \frac{\lambda_L L u_2}{\lambda_L} (e^{\lambda_L(t-T)} - 1) + \frac{\lambda_L L \alpha}{\lambda_L} (e^{\lambda_L(t-T)} - 1) \\
& - \frac{\sigma_X^2 u_1^2}{4\lambda_X} (e^{2\lambda_X(t-T)} - 1) - \frac{\sigma_L^2 \alpha^2}{4\lambda_X} (e^{2\lambda_X(t-T)} - 1) - \frac{\sigma_L^2 u_2}{4\lambda_L} (e^{2\lambda_L(t-T)} - 1) \\
& - \frac{\sigma_L^2 \alpha^2}{4\lambda_L} (e^{2\lambda_L(t-T)} - 1) - \frac{\sigma_L^2 \alpha u_2}{\lambda_X} (e^{\lambda_X(t-T)} - 1) + \frac{\sigma_L^2 u_2 \alpha}{2\lambda_L} (e^{2\lambda_L(t-T)} - 1) \\
& + \frac{\sigma_L^2 \alpha^2}{\lambda_X + \lambda_L} (e^{(\lambda_X + \lambda_L)(t-T)} - 1) - \frac{\rho \sigma_X \sigma_L \alpha u_1}{2\lambda_X} (e^{\lambda_X(t-T)} - 1) \\
& - \frac{\rho \sigma_X \sigma_L \alpha u_1 u_2}{\lambda_X + \lambda_L} (e^{(\lambda_X + \lambda_L)(t-T)} - 1) + \frac{\rho \sigma_X \sigma_L \alpha u_1}{\lambda_X + \lambda_L} (e^{(\lambda_X + \lambda_L)(t-T)} - 1).
\end{aligned}$$

Now, taking $(u_1, u_2) = (1, 0)$ and defining $m = -\lambda_X/(\lambda_X - \lambda_L)$ we obtain

$$\begin{aligned}
 M(t, T) = & m_1(e^{\lambda_X(t-T)} - 1) + m_2(e^{\lambda_L(t-T)} - 1) + m_3(e^{2\lambda_X(t-T)} - 1) \\
 & + m_4(e^{(\lambda_X + \lambda_L)(t-T)} - 1) + m_5(e^{2\lambda_L(t-T)} - 1), \quad (\text{B-17})
 \end{aligned}$$

where

$$\begin{aligned}
 m_2 = Lm \quad m_4 = \frac{m^2 \sigma_L^2}{\lambda_X + \lambda_L} + \frac{m \rho \sigma_X \sigma_L}{\lambda_X + \lambda_L} \quad m_5 = -\frac{m^2 \sigma_L^2}{4\lambda_L} \\
 m_1 = -\frac{\lambda_L Lm}{\lambda_X} \quad m_3 = -\left(\frac{\sigma_X^2}{4\lambda_X} + \frac{\sigma_L^2 m^2}{4\lambda_X} + \frac{m \rho \sigma_X \sigma_L}{2\lambda_X} \right).
 \end{aligned}$$

We now turn to the case of the jump model. Here N_1 and N_2 are as before, but M has two more components due to the jumps. We can write $M(t) = M_c(t) + M_1(t)$, where $M_c(t)$ is given by (B-17) and $M_1(t)$ satisfies:

$$\frac{dM_1}{dt} := \frac{N_1}{\eta_u - N_1}, \quad M_1(u, T, T) = 0. \quad (\text{B-18})$$

Integrating both sides we have

$$M_1 = \int \frac{u_1 e^{\lambda_X(t-T)}}{\eta_u - u_1 e^{\lambda_X(t-T)}} dt \quad (\text{B-19})$$

$$= \int \frac{1}{\frac{\eta_u e^{\lambda_X T}}{u_1} e^{-\lambda_X t} - 1} dt. \quad (\text{B-20})$$

Applying the formula

$$\int \frac{dx}{a + be^{cx}} = \frac{1}{ac} [cx - \ln(a + be^{cx})], \quad (\text{B-21})$$

with $a = -1$, $b = (\eta_u e^{\lambda_X T})/u_1$ and $c = -\lambda_X$ we get

$$M_1 = \frac{1}{\lambda_X} [-\lambda_X t - \ln(-1 + ((\eta_u e^{\lambda_X T})/u_1)e^{-\lambda_X t})] + C. \quad (\text{B-22})$$

Since $M_1(\mathbf{u}, T, T) = 0$,

$$C = T + \frac{1}{\lambda_X} \ln(-1 + \frac{\eta_u}{u_1}). \quad (\text{B-23})$$

Therefore

$$\begin{aligned} M_1 &= -(t - T) - \frac{1}{\lambda_X} \ln\left(-1 + \frac{\eta_u}{u_1 e^{\lambda_X(t-T)}}\right) + \frac{1}{\lambda_X} \ln\left(-1 + \frac{\eta_u}{u_1}\right) \\ &= -(t - T) - \frac{1}{\lambda_X} \left(\ln\left(\frac{-u_1 e^{\lambda_X(t-T)} + \eta_u}{u_1 e^{\lambda_X(t-T)}}\right) - \ln\left(\frac{-u_1 + \eta_u}{u_1}\right) \right) \\ &= -\frac{1}{\lambda_X} \ln\left(\frac{\eta_u - u_1 e^{\lambda_X(t-T)}}{\eta_u - u_1}\right). \end{aligned}$$