# Modelling Power Spikes with Inhomogeneous Markov-Switching Models

Vicke A. Norén

Master's Thesis
2013:E66

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

**Abstract**

Interest in modelling electricity prices has, despite its relatively short history, resulted in widespread types of models that tend to be too intricate to incorporate most price characteristics. This thesis pursues a flexible approach that comprehends stylized facts of electricity prices while it still handles complexity in order to facilitate calibration and forecasting. Although time-varying transitions of non-linear Markov-switching models add a new dimension to the problem, the extension is pivotal to encompass the timing of power spikes. Simulation studies provide a comparison between the maximum likelihood estimator and the EM algorithm and validate the precision of the estimators. A comprehensive study of the model framework in the independent regime setting that is applied to real data from the German and Nordic markets confirms the hypothesis that extensive models with exogenous variables outperform time-invariant counterparts. Improvements of electricity price dynamics and other issues involved in the process of modelling electricity prices as well as potential future research topics are also suggested and discussed.

**Acknowledgements**

I would like to thank my supervisor, Erik Lindström at the Centre for Mathematical Sciences at Lund University, for guidance and many insightful discussions.

# Contents

Chapter 1

# Introduction

## 1.1 Historical Background

The discoveries of electric phenomena that have led to large-scale electricity generation are a milestone in the human history. Ever since the industrialization took off, energy has served as the most important force driving the global development. Today electricity is indispensable for a steadily growing part of the population due to groundbreaking innovations in widespread areas. Imagine an everyday life without electricity and you soon realize how reliant we are on it. The role it has become for today's digitized society makes us take it for granted and therefore it has been given an exclusive treatment for a relatively long time.

Power generation has traditionally been centralized regulated with governments setting the agenda for designing power generation systems. The regulation imposed a vertical market structure and ensured production and supply to be secure and efficient in the sense that production plans should cover expected consumption. This tradition started to unravel in Chile in the early 1980s, whereupon many countries followed mostly due to ideological and political reasons as well as technological ones. While many asset classes, including energy resources such as crude oil, coal etc., were brought to markets for a relatively long time ago, energy output has been in a firm grip by the authorities. Electricity market liberalizations have opened up new markets and allowed contracts and trades to be settled for energy output as is rather than underlying energy resources. Whether the deregulation is flawed or not is a question outside the scope of this thesis, but the competitive markets that have replaced the initially monopolistic counterparts have facilitated trading in new markets for new participants. Nowadays a diversity of contracts can be bought and sold on electricity markets. Everything from standardized spot contracts to derivatives is traded, and for special purposes bilateral contracts can be settled on the over-the-counter market. For an introduction to everything from electricity market liberalization to a statistical modelling approach to study such markets see Weron (2006).

Energy markets are very interesting because of their multilateralism, which means that different methods have to be used depending on what behaviour one would like to

study. Fundamental models are preferable if one would like to gain greater insight into the markets and the systems with suppliers and consumers. For electricity markets in particular, this means that structural changes such as changes in power generation and capacity as well as outage and maintenance have to be described in order to plan for and draw conclusions from long-term scenarios.

A different approach that analyzes strategic decisions by market participants is based upon game theory and aims to understand what long-term behaviours are optimal in the Cournot-Nash framework. Another equilibrium approach is in addition the supply function equilibrium framework.

Statistical models have also drawn a lot of attention. While the former models try to understand long-term behaviour, the latter focus on short-term forecasts with applications in risk management and derivatives pricing, to name a few. Although this thesis is devoted solely to the statistical approach, it can be of interest to anyone that is trying to understand the dynamics that affect short-term electricity prices, since fundamental information works as determinants in many models. While the literature has mostly deployed the model framework to study short-term scenarios, the introduction of exogenous dependence in this thesis enables a deeper understanding of price sensitivity and its correlation with seasonal patterns. Some stylized facts of electricity prices shall, however, first be presented before proceeding to the statistical modelling.

## 1.2 Characteristics of Electricity Prices

Technology has driven the development of market structures insomuch that trades are for instance much more frequent, resulting in more liquid markets. What makes electricity price series highly interesting is that they are being "continuously" traded in the markets during the whole year. The highest volumes are represented by auction-based price settlements, whereafter continuous intraday trading serves to rebalance such prices. Day-ahead prices are technically the result of aggregate bid and ask prices for delivery in 24 hours. These are often settled on an hourly basis based upon the market equilibrium.

The fact that electricity has more idiosyncrasies than most asset classes, indeed, makes it a one of a kind asset class. Even though some characteristics are more obvious than others, most of them remain inevitable when modelling day-ahead spot prices; hence models have to incorporate the majority of the characteristics in order to be successful. For this reason the most important attributes will be presented for the rest of this section.

Like commodities, electricity prices are determined by supply and demand — more precisely by the equilibrium of the two — such that any shift in either supply or demand, or both of them, will have immediate consequences for the prices. Although electricity prices have some similarities to commodities as well as equity and bonds, it has to be treated in a quite different way. First and foremost it is very cost-ineffective to store electricity. In contrast to most commodities, non-storability limits the number of strategies market participants can take and implies that the risk-neutral measure is not unique. In addition to storing electricity, it is also difficult to transport electricity over long distances due to transmission constrains. These two factors imply that arbitrage opportunities in

electricity markets are very rare.

The relatively small number of market participants in electricity markets enables the possibility of influencing a market to a great extent. Power generation, in particular, tends to be concentrated to only a few suppliers. This together with market design forms the market. Not to mention, political decisions on important market mechanisms, such as tax levels, for the market per se or any closely related industry can of course disturb current market conditions and have substantial long-term effects. On the other hand, such interventions can cause economic incentives to decrease pollution and emission, and lead to development of new and refined types of energy production.

The idea of market design varies across countries. Some areas have imposed restrictions and subsidies in order to favour certain types of power generation, e.g. emission allowance and particularly carbon emission trading in the EU targeting carbon dioxide. Renewable energy has in this way experienced a vast growth over a relatively short time period. Injection of renewable energy into the energy markets has made power generation more unpredictable because of the difficulty in predicting production from such sources. While hydropower is not strictly tied to prediction of short-term perception, other renewable generations such as wind and solar power are to a much higher degree dependent on auspicious weather conditions. The main difference lies in hydropower being storable and therefore easier to moderate, whereas the latter are examples of highly uncertain energy sources as they might only last momentarily.

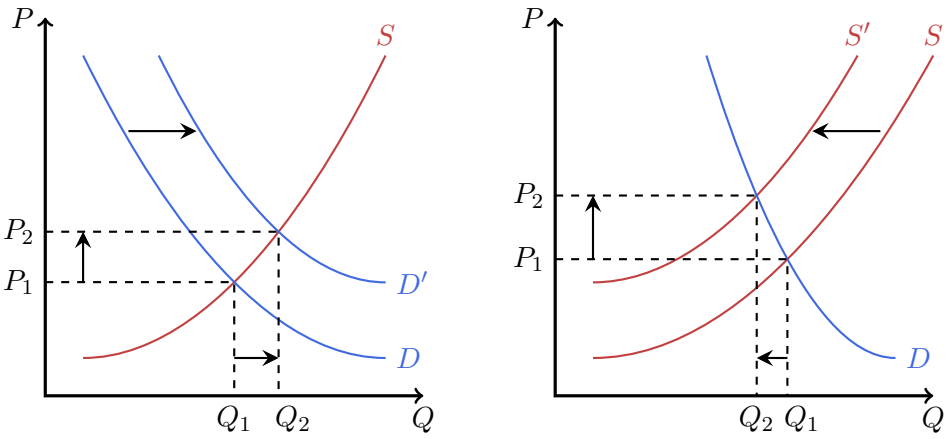
The fundamental idea here is that different energy sources affect generation, and consequently spot prices, in various ways. Intuitively, the more predominating a type of power generation is, the greater is its effect on prices overall and in particular the mean price level. The composition of power production shall, as a consequence, affect the volatility of price dynamics such that it is in proportion to the volatility of the power generation.

The supply curve is based upon predominating power generation and differs between countries. The supply is capped by the installed capacity for all generation units. The market network limits the quantity that can be transferred between areas and countries, thus flows of import and export of electricity are constrained to the capacities of the grids, which means that integration with other markets together with outages can cause a temporary surplus in demand.

Power plants and nuclear reactors are in general steady power sources and serve as a buffer so that there is some sort of limit to how severe changes in supply can become, although unexpected outage and activation of nuclear reactors and hydroelectric power stations take time to in order to benefit from them. Renewable energy has, by contrast, played an unobtrusive role in the history of power generation and still only accounts for a fraction of the world's total power production. Even though coal and fossils still are predominating, radical changes have occurred over the recent years. Renewable electricity production has grown considerably in some countries and this trend will most likely continue to hold on, albeit perhaps not at the same pace. Some regions have already been able to take advantage of some of their sustainable energy resources such as hydro, tidal, solar and wind power. These as well as other types of renewable sources have in

some cases become more substantial for the power generation. This paradigm shift has brought some uncertainty to the system as well. A situation where markets are more exposed to heavy price fluctuations has materialized due to the development of green energy, particularly in areas with a relatively high exposure to such kinds of resources, since a greater part of power generation coming from sustainable energy means higher correlation with external circumstances. Consequently, production planning has to be considered carefully in order to avoid mismatching.

A basic example of how price changes occur is illustrated in Figure 1.1. The left panel shows a typical change in demand that results in both higher price and quantity. A higher price also occurs if supply decreases, cf. the right panel. Notice the difference between the slopes of the demand curves; the demand curves in the left panel are not as steep as the fairly inelastic one in the right panel. Changes in supply or demand will therefore affect prices differently due to the price elasticity of demand. Electric energy consumption is considered very inelastic and abrupt changes in prices can thus easily occur, even if the supply level only changes slightly. The right panel in Figure 1.1 depicts a scenario where a sudden decrease in power generation, e.g. due to worsening weather conditions, causes a price hike. Notice the proportion of change in price to change in quantity.




Example of a shift in demand. Example of a shift in supply.

Figure 1.1: Changes in supply and demand curves in two different markets experiencing two schematically different scenarios that both result in higher prices. The left panel shows how a positive shift in demand pushes up market prices and results in higher produced quantity. In the right panel one can see how a negative shift in supply from $S$ to $S'$ affects prices heavily with the price difference $P_2 - P_1$ while the quantity difference $Q_2 - Q_1$ is relatively small due to a fairly inelastic demand. Both examples are typical for electricity markets and show how price spikes occur.

Another obvious reason for price changes is seasonal behaviour; prices follow seasonal trends during the year since seasonal conditions are intimately related to supply and demand, especially during summer and winter. Hot days tend to drive prices up since air

conditioning is needed; the same holds for cold days as heating is higher during such days. The supply side, on the other hand, can to some extent be reliant on e.g. wind, insolation and perception, of which the latter for instance affects reservoir levels in hydroelectric dams. Besides long-term trends, one can, in addition to holidays, also observe weekly cycles where weekends diverge from business days. A higher demand can be observed from Monday to Friday, since businesses and industries are more active during these days. In addition to weekly cycles, intraday behaviours are also apparent. Typical working days from morning to evening are coherent with higher demand, since industries run most of their work and activities during this time of the day. The intraday pattern is clear; surges in prices can be observed during morning hours followed by quite volatile behaviour all through the evening, during which soaring prices are evident once again. On the other hand, during the night the overall activity is low with few rapid price changes as a consequence.

Yet another property that is inherited by electricity prices is referred to as mean-reversion, which means that prices do deviate but tend to follow a particular price level over time. The deviation, which is partly deterministic and partly stochastic, is usually observed in so-called volatility clusters, i.e. time periods with similar volatility. Frequent occurrence of so-called spikes, i.e. price movements that differ remarkably from normal prices, can be observed during periods with high volatility. Positive spikes are commonly referred to as up-spikes or simply spikes, whereas negative spikes are referred to as down-spikes or drops. Such surges in spot prices characterize electricity markets simply because the amplitude of such movements can be several magnitudes greater than what has been observed in for instance commodity and stock markets.

Electricity and many commodity prices have also been observed to behave differently depending on the price level. While equity is known to display the so-called leverage effect, i.e. volatility tends to increase with lower prices, the effect is known to be opposite for electricity prices. Prices in the electricity market are instead subject to higher volatility along with higher prices, a phenomenon that is called the inverse leverage effect.

## 1.3 Overview of Statistical Modelling Approaches

Research in electricity markets has shown that, in order to construct robust models for forecasting short-term electricity prices, it is wise to incorporate the above-mentioned properties. Different approaches for modelling discrete-time price dynamics include autoregressive (AR) models, autoregressive-moving-average (ARMA) models and extensions to these with external input as well as integrated autoregressive-moving-average (ARIMA) models and seasonal autoregressive-moving-average (SARMA) models, or any combination thereof. Drift-diffusion processes have also shown to do well in continuous time, as have jump-diffusion processes.

The Nobel laureate Robert Engle suggested in Engle (1982) that volatility in econometrics is stochastic rather than flat, a proposition that later awarded him the Nobel prize. His seminal contribution to capture this behaviour lies in the usage of autoregressive conditional heteroskedasticity (ARCH) models. A few years later Bollerslev

(1986) extended this discrete-time model to incorporate lagged conditional variances in a generalized autoregressive conditional heteroskedasticity (GARCH) model. This way of modelling variability in financial data has proven to be very useful and ever since its introduction the popularity of conditional volatility models has increased tremendously.

Just like for discrete time, there are various stochastic volatility models in continuous time, see e.g. Musiela and Rutkowski (2005) for a collection of such models. However, such continuous-time models have not been applied to the same degree to energy markets as discrete-time models.

While approaches that include the family of GARCH models are interesting alternatives, models that involve regime-switching have become immensely popular in electricity markets and for this reason this thesis treats models with Markovian switching exclusively. These multifaceted models have also become proverbial for some applications and proven very successful in a vast number of areas, for example speech recognition (Rabiner (1989)), econometrics (Hamilton (1989)), bioinformatics (Krogh et al. (1994)), and environmental processes (Lu and Berliner (1999)), to name a few. An overview by Hamilton and Raj (2002), which is devoted to financial time series and especially business cycle durations, depicts the history of the model class and how it has developed along with its financial applications. A similar presentation of electricity price models is done by Janczura and Weron (2010), wherein comparisons of Markov-switching models show that many of the above-mentioned characteristics are actually captured by this model framework.

Weron et al. (2004) and de Jong (2006) among others apply and compare both jump-diffusion processes and models with regime-switching to electricity prices and suggest that the latter model class is preferable due to the way it distinguishes spikes. Weron (2006) discusses and compares a whole range of modelling approaches and highlights that Markov-switching models have a superior in-sample fit but at the same time have problems with out-of-sample forecasting. Further development of the model framework has since strengthened its already advantageous properties in order to perform better on both fronts. As an example, Erlwein et al. (2010) go one step further and model prices as a mixture of jump-diffusion processes with an underlying regime-switching mechanism determining the model parameters.

Despite the successes that the Markovian framework has brought to the electricity market, it is not necessarily the only model class to consider. Robinson (2000) shows that a non-linear logistic smooth transition autoregressive (LSTAR) model outperforms its linear counterpart. Surveys by Rambharat et al. (2005) and Misiorek et al. (2006) show that threshold autoregressive (TAR) and TARX models also perform well. The main shortcoming of TAR models is that their thresholds have to be explicitly stated in order to identify the output processes, while the model framework in this thesis does not require any a priori believes about the state transitions. Huisman et al. (2007) among others suggest to model prices separately for each hour, since volatility behaves differently during a day. In the case of Huisman et al. (2007) this is done in a panel framework. Yet another approach is taken by Wang and Ramsay (1998) and Szkuta et al. (1999), who model prices with artificial neural networks.

All the modelling approaches mentioned so far and dozens of others are regularly

tested by researchers, hence it is not the aim of this thesis to compare models outside the Markovian framework, which has received much attention in the modelling of electricity prices due to its many successes and therefore is of utmost interest.

## 1.4 Thesis Outline

Rather than modelling the volatility itself, this thesis chooses a route that has already proven successful in the literature, but which still lacks a strong connection between price series and fundamental information. The intuition is that such a connection ought to be prominent, yet few attempts have been made to justify its existence. State-of-the-art models are in this way implemented and developed in order to gain a better understanding of price dynamics and more accurate estimations. Before such models are described in detail, the theoretical framework is presented in a more general context since the underlying theory is not solely restricted to electricity prices.

The theoretic part with its emphasized discourse on Markov-switching models is followed by a thorough treatment of robust estimation techniques, viz. maximum likelihood and expectation-maximization. For the latter case closed-form expression are derived for almost all model parameters. Simulation studies then compare the accuracy of the estimation methods.

The classical two-state model suggested by Hamilton (1989) is still considered to be highly relevant, but is nevertheless quite limited. This thesis addresses the limitation of time-invariant transition probabilities by stressing the crucial importance seasonality has for calibrations and forecasts of prices. This pivotal extension of standard Markov-switching models gives rise to new problems. What fundamental information describes the underlying mechanics? Novel ingredients of such models are, in addition to bespoken indicators, examined exhaustively in order to try to determine the relationships between prices and determinants. It is worthwhile to stress that, while seasonality of electricity prices is well-known, few studies utilize this approach. This thesis can therefore serve as a guided tour of how to extended time-invariant Markov-switching models with respect to modelling as well as estimation aspects.

Electricity price models are then treated exclusively, especially the renowned models based upon the aforementioned theory. Two model frameworks are of particular interest; models with dependent and independent regimes.

The subsequent chapter is devoted to analysis of Nord Pool Spot (NPS), which comprises of the Nordic and Baltic countries, and the European Energy Exchange (EEX) in Germany — two of the largest electricity markets in the world. These two prominent markets stand out with interesting combinations of energy sources, see Figure 1.2, which compares the installed capacity as well as net electricity generation. Both areas have significant proportions of renewable energy, especially Germany which has just over one third of its installed capacity in unpredictable wind turbines and solar cells, which in turn ought to be substantial factors in volatile price movements. The figures do not reveal that the Nordic area is quite differentiated due to natural resource allocation. Above all, Norway relies heavily on hydropower, whereas Denmark has an even greater installed

capacity in wind farms than Germany.

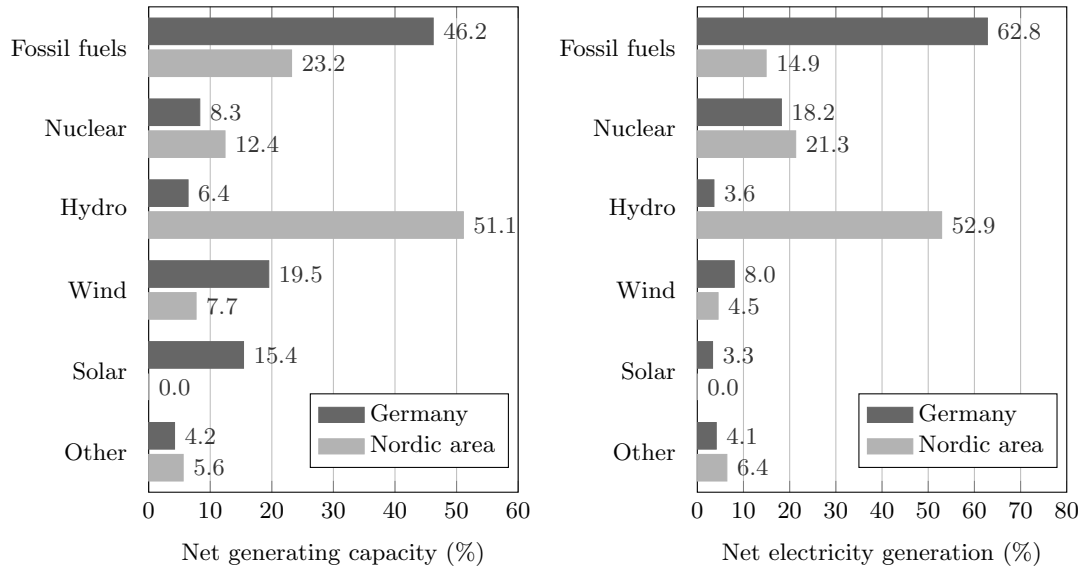



Figure 1.2: Comparison of net generating capacity and net electricity generation by source in Germany and the Nordic area[1] for 2011. Data according to the Statistical Yearbook 2011 from the European Network of Transmission System Operators for Electricity.

The analysis includes exogenous variables that are of special interest for each market. Such variables include consumption and production as well as wind power generation and forecasts thereof. An impending question is for example whether the wind power generation, particularly in Denmark, affects the system price. Keep in mind that Germany also has a differentiated energy portfolio. Could some sources have connections to abrupt price movements? The German market is predominated by solid fuels, whereas natural gases constitute the third greatest energy source, thus one could argue that emission allowances and gas prices affect electricity prices, the question is to what extent. This thesis will try to answer these questions among many other ones.

In summary, this thesis aims to shed light on electricity price movements, both their characteristics and what affects them. The main contribution of this thesis is the extension to time-varying transitions such that three-state models become non-linear in both state and space. In particular, the study of wind power generation is, to my best knowledge, novel in this model framework. A second contribution is the study of both daily and hourly prices. Simulations studies of an inhomogeneous three-state model also provide a statistical comparison of the accuracy of the estimation techniques.

[1] Includes Denmark, Finland, Norway and Sweden.

Chapter 2

# Price Dynamics and Markov-Switching Models

## 2.1 Stochastic Dynamics in Continuous Time

In order to model price dynamics in continuous time it is customary to use some sort of stochastic process. Consider for instance the linear (and possibly multidimensional) stochastic differential equation

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{2.1}$$

where $W(t)$ is a Wiener process, i.e. a standard Brownian motion. Besides having a certain drift $\mu$, which is possibly time-varying, the process is driven by its stochastic component $\sigma$. A lot of research has dealt with how to better capture some known behaviours of a certain process by expanding the class of stochastic processes. Stochastic differential equations have their origins in physics. In an early work, however, Bachelier (1900) suggested randomness to be applied to financial applications. It took some time until the ideas of Bachelier (1900) got realized. The class of stochastic processes grew in the meantime, during which Uhlenbeck and Ornstein (1930) introduced an important stochastic process, namely

$$dX(t) = \kappa(\vartheta - X(t))dt + \sigma dW(t),$$

which has become known as the Ornstein-Uhlenbeck process. By applying Itô's lemma to $df(t, X(t)) = e^{\kappa t}X(t)$ and then integrating with respect to $t$ the solution is readily obtained as

$$X(t) = X(0)e^{-\kappa t} + \vartheta(1 - e^{-\kappa t}) + \sigma \int_0^t e^{\kappa(s-t)}dW(s).$$

The long-term expectation of the solution is $\vartheta$ and is known as the long-term mean-reversion level, which means that the Ornstein-Uhlenbeck will revert to this level over time. Not only does the process possess mean-reversion, but it also explicitly specifies

the mean-reversion level, thus making it covetable for applications that exhibit such kind of behaviour. The speed of the reversion is in addition determined by $\kappa > 0$.

Vašíček (1977) proposed the Ornstein-Uhlenbeck process to model short-term interest rates. The (inverse) leverage effect was incorporated in the constant elasticity of volatility (CEV) model for option prices which was introduced by Cox (1975). That model, however, lacks mean-reverting characteristics. A model including both mean-reversion and the leverage effect was proposed by Chan et al. (1992). This extended non-linear model,

$$dX(t) = \kappa(\vartheta - X(t))dt + \sigma X(t)^{\gamma} dW(t), \tag{2.2}$$

is referred to as the CKLS model and is a local volatility model rather than a pure stochastic volatility model. Chan et al. (1992) conducted a profound investigation of a variety of short rate models and showed that the CKLS model performed well. It is essentially a generalization of a bunch of well-known short rate models. It equals the geometric Brownian motion (GBM) used in the classical Black and Scholes (1973) model when $\vartheta = 0$ and $\gamma = 1$. Moreover, when $\gamma = 0$ it boils down to the Vašíček (1977) model, when $\gamma = 1$ it coincides with the Brennan-Schwartz model by Brennan and Schwartz (1980), and when $\gamma = 1/2$ it equals the widely used Cox-Ingersoll-Ross (CIR) model proposed by Cox et al. (1985).

Besides the local volatility model used in this thesis, more extensive ways of specifying volatility is of course possible. By expanding the model to a two-factor model specifying both price dynamics and underlying volatility dynamics, the instantaneous volatility becomes a hidden process with a continuous state space, cf. the ARCH model coined by Engle (1982).

A different approach involving fractals, dubbed Markov-switching multifractal (MSM) models, was developed by Calvet and Fisher (2004). It has some similarities to the aforementioned models as it captures multiple frequency components in volatility. One of its strengths is that the number of latent states does not have to be specified in advance as this number is estimated as a part of the estimation process.

We shall henceforth focus exclusively on another model class known as Markov-switching models. Like MSM models, they also capture stochastic volatility by allowing for switches between different regimes. Since both the above-mentioned price dynamics and models with Markovian switching serve as corner stones for the rest of the thesis, an exhaustive treatment shall be carried out in the Markovian framework by explaining both model dynamics and transition mechanisms.

## 2.2 Markov Chains and Their Properties

This section lays the foundation of the theoretical framework of the thesis. The outline is set to provide the prerequisites in order to grasp the theory behind the modelling. Before proceeding with the core framework, let us first recall some basic notations in probability theory.

The space $\Omega$ of all possible outcomes is called sample space. A measurable space $(\Omega, \mathcal{F})$ is a sample space with a $\sigma$-algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$, hence a $\sigma$-algebra, whose elements

are called events, is a non-empty collection of subsets of $\Omega$ that, in addition, is closed under complements, countable unions and (by De Morgan's laws) countable intersections. A filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ is an indexed family of $\sigma$-algebras on $\Omega$ such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for all $s \leqslant t$. One can associate filtrations with increasing flows of information. A finite, non-negative measure $\mu$ on a measurable space is a probability measure, commonly denoted by $\mathbb{P}$, if it has unit total mass, that is $\mu(\Omega) = 1$. In such case the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is referred to as probability space. While the purist may insist, we write for the sake of notational convention $\mathbb{P}(\xi(\omega) \in E)$ throughout this thesis in lieu of $\mathbb{P}(\{\omega\colon \xi(\omega) \in E\})$.

Consider two measurable spaces $(\Omega, \mathcal{F})$ and $(\mathsf{X}, \mathcal{X})$. A function $\xi\colon \Omega \to \mathsf{X}$ is measurable with respect to $\mathcal{F}$ if $\xi^{-1}(E) \in \mathcal{F}$ for every $E \in \mathcal{X}$. Furthermore, consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. A function $\xi$ on $\Omega$ is adapted to the filtration $\{\mathcal{F}_t\}$ if $\xi_t \in \mathcal{F}_t$ for all $t \geqslant 0$.

Let $\{\Omega_i\}_{i\in I}$ be a family of sets indexed by a non-empty set $I$. The Cartesian product is defined by

$$\prod_{i\in I} \Omega_i = \{f\colon I \to \cup_{i\in I}\Omega_i \mid f(i) \in \Omega_i \text{ for all } i \in I\}.$$

Now let $\{(\Omega_i, \mathcal{F}_i)\}_{i\in I}$ be a family of measurable spaces indexed by a non-empty set $I$. The set of all measurable rectangles is defined by

$$\coprod_{i\in I} \mathcal{F}_i = \left\{ \prod_{i\in I} E_i\colon E_i \in \mathcal{F}_i \ , \ E_i \neq \Omega_i \text{ for finitely many } i \in I \right\}.$$

Finally, the product $\sigma$-algebra is defined as the (smallest) $\sigma$-algebra generated by all measurable rectangles, i.e.

$$\bigotimes_{i\in I} \mathcal{F}_i = \sigma\left( \coprod_{i\in I} \mathcal{F}_i \right).$$

Throughout this thesis bold notation emphasize vector notation (or matrix notation when appropriate) and in particular the notation $\mathbf{X}_t$ denotes the vector $(X_1, \ldots, X_t)$ representing the sequence $\{X_s\}_{s\in S}$, where $S = \{n \in \mathbb{N}\colon n \leqslant t\}$. Parameter estimates are, when appropriate, denoted by superscript to indicate the iterative step, whereas the circumflex symbol is reserved for final parameter estimates.

We are now in position to formulate what makes a stochastic process a Markov chain.

**Definition 2.1 (Markov Chain).** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $\mathbb{F}$ and let $(\mathsf{X}, \mathcal{X})$ be a measurable space. A stochastic process $\{X_t\}_{t\geqslant 0}$ on $\Omega$, where every random variable of the process is an $\mathcal{F}$-measurable function $X\colon \Omega \to \mathsf{X}$ that takes values in the state space $\mathsf{X}$, is a Markov chain under $\mathbb{P}$ if it is adapted to $\mathbb{F}$ for all $t \geqslant 0$ and, for all events $E \in \mathcal{X}$,*

$$\mathbb{P}(X_{t+1} \in E \mid \mathcal{F}_t) = \mathbb{P}(X_{t+1} \in E \mid X_t). \tag{2.3}$$

In other words, a Markov chain is a random process where the future outcome only depends on today's value, which basically means that the chain has no memory of its past. This restriction can be tweaked by allowing a Markov chain to depend on its past, in which case it is convenient to refer to the order of the chain; first-order Markov chains are commonly referred to as simply Markov chains, whereas $s$-order chains obey

$$\mathbb{P}(X_t \in E \mid \mathcal{F}_{t-1}) = \mathbb{P}(X_t \in E \mid X_{t-1}, \dots, X_{t-s}).$$

In this context it is convenient to relate Markov chains to transition kernels.

**Definition 2.2 (Transition Kernel).** *Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be two measurable spaces. A function $Q\colon \mathsf{X} \times \mathcal{Y} \to [0, \infty]$ is a transition kernel if it satisfies*

*(i) for all $x \in \mathsf{X}$, $Q(x, \cdot)$ is a positive measure on $(\mathsf{Y}, \mathcal{Y})$;*

*(ii) for all $E \in \mathcal{Y}$, the mapping $\mathsf{X} \ni x \mapsto Q(x, E)$ is measurable.*

*Moreover, $Q$ is a Markov transition kernel if $\mathsf{X} = \mathsf{Y}$ and if $Q(x, \mathsf{X})$ is a probability measure for all $x \in \mathsf{X}$.*

From the definition it is clear that the kernel $Q$ acts both as a function $Q(\cdot, E)$ as well as a measure $Q(x, \cdot)$ and that a Markov chain is governed by a Markov transition kernel $Q\colon \mathsf{X} \times \mathcal{X} \to [0, 1]$ such that

$$Q_t(X_t, E) = \mathbb{P}(X_{t+1} \in E \mid X_t),$$

for all states $X \in \mathsf{X}$ and events $E \in \mathcal{X}$. The relationship between a transition kernel and a transition density $q$ is given by

$$Q_t(X_t, E) = \int_E q_t(x_t, dy).$$

In the case of a countable state space the transition kernel can be interpreted as a transition probability matrix $\mathbf{P}$, i.e. a right stochastic $\mathsf{X} \times \mathsf{X}$ matrix with the transition probability $p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i)$ denoting the probability of moving from state $i$ to state $j$ on the corresponding index. From this notation it is clear that the constraints

$$\begin{aligned} 0 \leqslant p_{ij} \leqslant 1, &\quad \text{for all } (i, j) \in \mathsf{X}^2, \\ \sum_{j \in \mathsf{X}} p_{ij} = 1, &\quad \text{for all } i \in \mathsf{X}, \end{aligned}$$

are jointly satisfied as direct consequences of the Kolmogorov axioms.

A Markov chain is time-homogeneous if its transition kernel does not depend on time, which implies that the Markov property (2.3) does not depend on $t$, otherwise the chain is inhomogeneous. Furthermore, a Markov chain is said to be stationary if it can reach a distribution that does not change over time. Such a distribution, $\boldsymbol{\pi}$, is called a stationary distribution of the chain and implies that the global balance equation

$$\int_{\mathsf{X}} \pi(x) q(x, y) dx = \pi(y) \tag{2.4}$$

holds for all $y \in \mathsf{X}$. For a countable state space, the global balance equation (2.4) can be written as $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, which is an eigenvalue equation with eigenvector $\boldsymbol{\pi}$ and unit eigenvalue.

Far from all Markov chains have a stationary distribution and unless some conditions are imposed on a chain it may have several stationary distributions or none at all. More precisely, it can be shown that an irreducible Markov chain has a unique stationary distribution if and only if all its states are positive recurrent. With irreducible is meant that all states communicate with each other, that is there exists an integer $n \in \mathbb{N}$ such that the $n$-step transition probability $p_{ij}^{(n)} = \mathbb{P}(X_{t+n} = j \mid X_t = i)$ is positive for all $(i, j) \in \mathsf{X}^2$ and for all $t$, whereas a recurrent state is a state that is $\mathbb{P}$-almost surely revisited by the chain, i.e. $\mathbb{P}(X_t = i \mid X_0 = i) = 1$ for some $t > 0$. In addition, positive recurrent means that the expectation of the recurrence time

$$\tau_i = \inf\{t \geqslant 1 \colon X_0 = i, X_t = i\}$$

of state $i$ is finite. All states of an irreducible Markov chain with finite state space are in fact positive recurrent.

The distribution of $X_t$ is simply denoted by the unconditional probability $\mathbb{P}(X_t = i)$ and the distribution of $X_0$ is in particular denoted by $\rho$ and referred to as the initial distribution of the chain. The period of state $i$ is the greatest common divisor $d$ of all $n$ such that $\mathbb{P}(X_{t+n} = i \mid X_t = i) > 0$; in particular, a state is aperiodic if $d = 1$. A simpler definition of aperiodicity is to require that

$$\int_{\mathsf{X}} Q^n(x, x)dx > 0$$

for sufficiently large $n$ in order to ensure the chain to be aperiodic, whereas for a finite state space this translates into a positive trace of $\mathbf{P}$. Furthermore, a Markov chain is ergodic if it is irreducible, positive recurrent and aperiodic. The following result shows that an ergodic Markov chain converges to a unique stationary distribution independently of its initial distribution.

**Theorem 2.3 (Limit Theorem).** *Let $\{X_t\}_{t\geqslant 0}$ be an ergodic Markov chain with state space $\mathsf{X}$. For all $x \in \mathsf{X}$ it then holds that*

$$p(X_t = x) \to \pi(x) \quad \text{as} \quad t \to \infty,$$

*independently of how the initial distribution $\rho$ is chosen.*

*Proof.* For an ingenious coupling argument see for instance Norris (1998). □

**Remark 2.4** *An important result of Erdős et al. (1949) shows that the limit of the sequence $\{p_{ij}^{(n)}\}$ of an ergodic Markov chain has a unique stationary distribution that is given by Kac's Theorem*[1] *as precisely $\pi(x) = 1/\mathbb{E}_x(\tau_x)$. With this in mind, a different*

[1] For a generalization of Kac's Theorem see Theorem 6.37 in Robert and Casella (2004).

*approach of the proof showing existence and uniqueness by applying Fatou's lemma (in the case of an infinite state space) and Lebesgue's dominated convergence theorem*[2] *is demonstrated by Brzeźniak and Zastawniak (2002).*

Another important result in the theory of Markov chains is the Perron-Frobenius theorem which states that an irreducible Markov chain with finite state space and period $d$ has at least one eigenvalue $\lambda = 1$, thus satisfying the global balance equation and guaranteeing that the transition probability matrix $\mathbf{P}$ can be decomposed into a corresponding eigenvector equal to the desired stationary distribution. Moreover, $\mathbf{P}$ has $d$ eigenvalues that satisfy $|\lambda| = 1$, while the remaining eigenvalues satisfy $|\lambda| < 1$. Indeed, the case of an ergodic Markov chain ensures that there exists only one $\lambda = 1$ and that the global balance equation has a unique solution. Consequently, decomposing $\mathbf{P}$ into the eigenvector corresponding to $\lambda = 1$ will ensure that the eigenvector equals the stationary distribution.

The stationary distribution of an ergodic Markov chain can be obtained directly by following Hamilton (1994) and setting

$$\mathbf{A}\boldsymbol{\pi} = \mathbf{e}_{K+1},$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} - \mathbf{P} \\ \mathbf{1} \end{pmatrix},$$

$\mathbf{e}_{K+1}$ is the $(K+1)$th column of the identity matrix $\mathbf{I}_{K+1}$ and $\mathbf{1}$ is a row vector of ones. As in Hamilton (1994), the solution is then given by

$$\boldsymbol{\pi} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}_{K+1}, \tag{2.5}$$

where $\mathbf{A}^T$ denotes the matrix transpose of $\mathbf{A}$.

### 2.2.1 Inhomogeneous Markov Chains

A Markov chain is time-homogeneous if it is time-homogeneous for all times $t$, cf. equation (2.3), otherwise it is called inhomogeneous as it possesses a time-varying behaviour, i.e. the Markov property (2.3) does indeed depend on time $t$ for some $t > 0$. There are numerous ways to let transition probabilities vary over time. Any mapping $\mathbb{R}^d \mapsto [0, 1]$ for any positive integer $d$ will make sure that the transition probabilities are valid probabilities as required. Incorporating exogenous processes into such a mapping is of course not only possible but also very tempting. Such methodological contributions date back to Goldfeld and Quandt (1973), who used Gaussian cumulative distribution functions for the regime-switching mechanism. To make sure proper inference still can be conducted, the surjective functions $f\colon \mathbb{R}^d \times \Theta \to [0, 1]$, where $\Theta \subset \mathbb{R}^d$, that govern the transitions shall be considered given that they can be constructed in some feasible ways.

[2] See for instance the theorems 11.31 and 11.32 in Rudin (1976).

In order to comply with the above reasoning and still retain simplicity, a logistic function of the form

$$f(\mathbf{z}, \boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}\cdot\mathbf{z}}}{1 + e^{\boldsymbol{\beta}\cdot\mathbf{z}}} \tag{2.6}$$

shall serve our purposes as it fulfills the extensive requirements. The concise vector notation reveals that in this way the transitions are governed by any $d$-dimensional vectors $\boldsymbol{\beta}$ and $\mathbf{z}$; the interesting choice here is, in particular, to consider exogenous information $\mathbf{z}$ with weights $\boldsymbol{\beta}$. The logistic function is tractable for capturing explanatory information, whereupon standard statistical test can be readily conducted in order to draw conclusions from the exogenous data.

Transition probabilities with logistic functions of the form (2.6) only obey Kolmogorov's axioms when $p_{ij} > 0$ for precisely two states $j \in \mathsf{X}$ and $p_{ik} = 0$ for all $k \neq j$. This limitation is a bit cumbersome in the case of more general state spaces. An extension to equation (2.6) is the multinomial modification

$$f(\mathbf{z}, \boldsymbol{\beta}, m) = \frac{e^{\boldsymbol{\beta}_m\cdot\mathbf{z}_m}}{1 + \sum_{n=1}^{N} e^{\boldsymbol{\beta}_n\cdot\mathbf{z}_n}},$$

which ensures $p_{ij} \in [0, 1]$ for all $(i, j) \in \mathsf{X}^2$, regardless of the number of states. Note that, while the above notation does not necessary imply time-dependence, a slight abuse of notation is used when letting $\mathbf{z}$ be time-dependent. Rather than the notational convention set for this thesis, we let $\mathbf{z}_t$ be a vector with $d$ observations at time $t$, i.e. $\mathbf{z}_t = (z_{1,t}, \ldots, z_{d,t})$, and $\mathbf{z}_{m,t}$ denote a contiguous subsequence of $\mathbf{z}_t = (\mathbf{z}_{1,t}, \ldots, \mathbf{z}_{n,t})$ such that $1 \leqslant m \leqslant n$. Similarly, $\boldsymbol{\beta}_m$ is a contiguous subsequence of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)$.

Logistic regression is a natural way of governing transitions, in part due to its resemblance to linear regression, and many models in the literature tend to exercise this option, although there are examples of other approaches. Logistic functions are for instance suggested by Diebold et al. (1994) and Zucchini and MacDonald (2009) among others. Zucchini and MacDonald (2009) admit that cases of more than two states are complicated, albeit surmountable. For an example with logistic mixture models as a valid alternative see Wong and Li (2001).

## 2.3 Markov-Switching Models

A state-space model, $\{(X_t, Y_t)\}_{t\geqslant 0}$, is a bivariate process that is representing an observable, $\mathsf{Y}$-valued process $\{Y_t\}_{t\geqslant 0}$ that is driven by a hidden process $\{X_t\}_{t\geqslant 0}$ that takes values in $\mathsf{X}$. Under the assumption that $X$ is a Markov chain, the general model becomes

$$\begin{aligned} X_{t+1} \mid X_t = x_t &\sim Q_t(x_t, \cdot), \\ Y_t \mid X_t = x_t &\sim f(x_t, \cdot), \end{aligned} \tag{2.7}$$

where the upper equation of (2.7) is the state equation (or evolution equation) of the underlying chain governed by the transition kernel $Q$ and the lower equation is the observation equation specified by observation (or emission) distribution $f$.

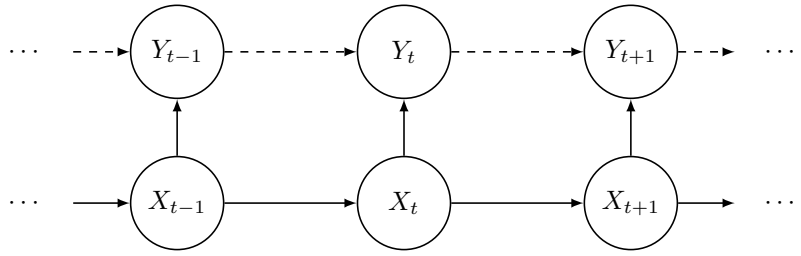



Figure 2.1: A Markov-switching model represented by a directed acyclic graph where $X$ is the latent process and $Y$ is the observable process. The model is a special case of a hidden Markov model when $Y$ has no autoregressive components.

A particular class of state-space models are hidden Markov models (HMMs), which require the state space of a Markov chain to be countable, hence the transition kernel can be interpreted as a (right) stochastic $\mathsf{X} \times \mathsf{X}$ matrix. The observations given a certain state are conditionally independent and the conditional distribution of $Y_n$ depends only on $X_n$.

Markov regime-switching (MRS) models, which are commonly referred to as Markov-switching models or sometimes Markov jump systems, are a generalization of hidden Markov models with the slight difference that the observation distribution in equation (2.7) allows lagged dependencies of the output process instead of observations being conditionally independent given the state process, see Figure 2.1 for a graphical representation. Due to the great importance of this model framework, a more formal definition is in order.

**Definition 2.5 (Markov-Switching Model).** *Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be two measurable spaces and let $Q$ be a Markov transition kernel on $(\mathsf{X}, \mathcal{X})$. Furthermore, let $G$ be a transition kernel from $(\mathsf{X} \times \mathsf{Y}^d, \mathcal{X} \otimes \mathcal{Y}^{\otimes d})$ to $(\mathsf{Y}, \mathcal{Y})$. The bivariate process $\{(X_t, Y_t)\}_{t \geqslant 0}$ with the transition kernel*

$$H_t[(x, y), E] = \iint_E Q_t(x, dx') G_t[(x', y), dy'],$$

*for all $(x, y) \in \mathsf{X} \times \mathsf{Y}^d$ and $E \in \mathcal{X} \otimes \mathcal{Y}$, is called a Markov-switching model.*

Any observation $y$ is, by the definition, allowed to influence future outcomes, cf. Figure 2.1. In this way the observation distribution in equation (2.7) is extended to $f(x_t, \tilde{\mathbf{y}}_{t-1}, \cdot)$, where $\tilde{\mathbf{y}}_{t-1}$ is a subsequence of $\mathbf{y}_{t-1}$, in order to include lagged values of $y_t$. Markov-switching models coincide with hidden Markov models if $G \colon \mathsf{X} \to \mathsf{Y}$, which in terms of the emission distribution means that $f(x_t, \mathbf{y}_{t-1}, \cdot) = f(x_t, \cdot)$ for all times $t$. Although MRS models are more complex than HMMs, they still share some nice properties making them very interesting from a modelling perspective while preserving the utilization of similar estimation techniques that are used for statistical inference in hidden Markov models.

If the underlying Markov chain of an MRS model is time-homogeneous, then we refer to the Markov-switching model as time-homogeneous, otherwise as inhomogeneous.

Chapter 3

# Estimation Methods in a Markov-Switching Framework

The objective of state-space modelling is to compute optimal estimates of hidden states given observed data. Both maximum likelihood and expectation-maximization, which are presented and discussed herein, are canonical methods of inference in Markovian frameworks.

## 3.1 Maximum Likelihood

An intuitively appealing idea is to try to describe data with some parameterized function and then find the parameter values that best describe the data. This reasoning leads us to the maximum likelihood (ML) approach. Suppose that the function $\mathcal{L}$ belongs to a certain family of distributions, i.e. $\mathcal{L} \in \{\mathcal{L}(\cdot\,;\boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\}$, where $\Theta \subset \mathbb{R}^d$ for some integer $d$ and $\boldsymbol{\theta}$ is a parameter vector for this family. Furthermore, assume that we have $T$ observations $y_1, \ldots, y_T$, thus we can write the likelihood function as

$$\mathcal{L}(y_1, y_2, \ldots, y_T\,;\theta) = p(y_1, y_2, \ldots, y_T \mid \theta).$$

We wish to find the estimate $\hat{\theta}$ that best approximates the sample, hence we would like to maximize the likelihood function. The maximum likelihood estimator (MLE) is defined by

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(y_1, \ldots, y_T\,;\theta).$$

This optimization problem can be rather hard to solve and therefore it is convenient to make use of the log-likelihood function $\ell$, defined by

$$\ell(y_1, \ldots, y_T\,;\theta) = \log \mathcal{L}(y_1, \ldots, y_T\,;\theta), \tag{3.1}$$

as the problem stays intact with respect to this transformation since the logarithm is a monotonic function.

### 3.1.1 Statistical Properties of the Maximum Likelihood Estimator

Before proceeding with the inference, let us for a moment review the MLE in a brief intermezzo. Maximum likelihood estimation is a sound approach to follow, not only due to its intuitive nature but also for its properties, of which three are very appealing. First, the MLE is consistent, which loosely speaking means that a sequence of maximum likelihood estimators converges to the true parameter value $\theta_\star$ as the sample size increases. Under rather mild conditions, especially the parameter space $\Theta$ is required to be compact and the log-likelihood function $\ell(\cdot\,;\theta)$ is required to be continuous in $\theta$ and bounded from above, it can be shown that strong consistency holds, i.e. $\hat{\theta} \xrightarrow{a.s.} \theta^\star$, since a sequence of MLEs converges almost surely to the true parameter value, i.e.

$$\sup_{\theta\in\Theta} |n^{-1}\ell(\mathbf{y}_n\,;\theta) - \ell(\theta)| \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty,$$

where $\ell(\theta)$ is a continuous deterministic function with a unique global maximum at $\theta_\star$. Almost sure convergence implies convergence in probability, which is enough for weak consistency. Second, provided that the assumptions for strong consistency hold and that, furthermore, $\theta$ is an interior point of $\Theta$, $\mathcal{L}(Y\,;\theta) \in \mathcal{C}^{d,2}$ for any (non-negative) integer $d$, and the Fisher information, defined by

$$\mathcal{I}(\theta) = \mathbb{E}\left[\left(\frac{\partial \log \mathcal{L}(Y\,;\theta)}{\partial\theta}\right)^2\right],$$

is non-singular, then the MLE is asymptotically normal, that is the following convergence in distribution holds:

$$\sqrt{n}(\hat{\theta}_n - \theta_\star) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\theta_\star)^{-1}) \quad \text{as} \quad n \to \infty,$$

where $\hat{\theta}_n = \arg\max_{\theta\in\Theta} \ell(\mathbf{y}_n\,;\theta)$. Note that in the case of a finite parameter vector of length $p$, the Fisher information takes the form of an $p \times p$ Hessian matrix where every element is

$$\mathcal{I}(\theta)_{i,j} = -\mathbb{E}\left(\frac{\partial^2 \log \mathcal{L}(Y\,;\theta)}{\partial\theta_i \partial\theta_j}\right).$$

Finally, what makes maximum likelihood estimation so covetable is its ability to attain the Cramér-Rao lower bound

$$\mathcal{I}(\theta)^{-1} \leqslant \mathbb{V}(\theta),$$

thus making the MLE an efficient estimator. In other words, if there exists an unbiased minimum variance estimation, then maximum likelihood will produce it. The Cramér-Rao lower bound, or any lower bound for that matter, is helpful in determining whether a problem is feasible or not by explicitly stating the minimum variance that then is comparable to a desired estimation precision. For a rigorous treatment of the properties of the maximum likelihood estimator, see for instance Cappé et al. (2005).

### 3.1.2 Filtering and the Forward Recursion

Recall a Markov-switching model with the latent process $\{X_t\}_{t=1}^T$ and the observable process $\{Y_t\}_{t=1}^T$ with state spaces $\mathsf{X}$ and $\mathsf{Y}$, respectively. The aim is to perform inference on the former process given the observations $y_1, \ldots, y_T$. Through the chain rule one sees that it is possible to write the incomplete log-likelihood function (3.1) as

$$\ell(\mathbf{y}_T\,;\boldsymbol{\theta}) = \sum_{t=1}^{T} \log p(y_t \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}). \tag{3.2}$$

By applying the law of total probability to equation (3.2) we get

$$\sum_{t=1}^{T} \log p(y_t \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}) = \sum_{t=1}^{T} \log \int_{\mathsf{X}} p(y_t \mid x_t, \mathbf{y}_{t-1}, \boldsymbol{\theta}) p(x_t \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}) dx_t, \tag{3.3}$$

and by finally applying the law of total probability and Bayes' theorem to equation (3.3) we arrive at

$$\begin{aligned}
&\sum_{t=1}^{T} \log \int_{\mathsf{X}} p(y_t \mid x_t, \mathbf{y}_{t-1}, \boldsymbol{\theta}) \int_{\mathsf{X}} p(x_t \mid x_{t-1}) p(x_{t-1} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}) dx_{t-1} dx_t \\
&= \sum_{t=1}^{T} \log \int_{\mathsf{X}} p(y_t \mid x_t, \mathbf{y}_{t-1}, \boldsymbol{\theta}) \\
&\qquad \times \int_{\mathsf{X}} \frac{p(x_t \mid x_{t-1}) p(y_{t-1} \mid x_{t-1}, \mathbf{y}_{t-2}, \boldsymbol{\theta}) p(x_{t-1} \mid \mathbf{y}_{t-2}, \boldsymbol{\theta})}{p(y_{t-1} \mid \mathbf{y}_{t-2}, \boldsymbol{\theta})} dx_{t-1} dx_t.
\end{aligned} \tag{3.4}$$

This deduction has lead to a recursive algorithm. Besides observing that the filtering distribution $p(x_{t-1} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta})$ in equation (3.3) induces the prediction probability $p(x_{t-1} \mid \mathbf{y}_{t-2}, \boldsymbol{\theta})$, note that, for the denominator in equation (3.4), one can make use of the same rewriting of the marginal likelihood as in equation (3.3) in order to efficiently implement this algorithm. Filtering distributions are then readily obtained by simply normalizing prediction distributions such that

$$p(x_t \mid \mathbf{y}_t) = \frac{p(x_t \mid \mathbf{y}_{t-1}) p(y_t \mid x_t, \mathbf{y}_{t-1})}{\int_{\mathsf{X}} p(x_t \mid \mathbf{y}_{t-1}) p(\mathbf{y}_t \mid x_t, \mathbf{y}_{t-1}) dx_t}.$$

Numerical optimization can be implemented once the filtering problem has been solved, for instance with a quasi-Newton method or Nelder-Mead simplex search method. Maximum likelihood estimation does not, however, implicitly assume that parameter constraints are satisfied, thus enabling for instance transition probabilities outside the unit interval. Unless the obtained MLE is consistent with a priori parameter bounds, putting constraints on at least some parameters is a wise choice in order to prevent erroneous parameters estimates. More specifically, transition probabilities are to be bounded to the unit interval and any variance terms are to be bounded to the non-negative real numbers.

## 3.2 Expectation-Maximization

An algorithm frequently used in state-space modelling is the expectation-maximization (EM) algorithm, see Algorithm 3.1. It was first proposed in a landmark paper by Baum et al. (1970) and later brought to its more general form by Dempster et al. (1977). It is essentially based upon two steps; the expectation step (E-step) and the maximization step (M-step). In the first step the expectation of the complete log-likelihood function $\log \mathcal{L}(\mathbf{X}, \mathbf{Y}\,; \boldsymbol{\theta})$ is computed, and in the second step the intermediate quantity $\mathcal{Q}(\theta\,; \theta^{(i-1)})$ is maximized given the optimal parameter estimate $\theta^{(i-1)}$ obtained at the previous iteration. Since the last updated parameters are the most optimal parameters obtained so far, it can be shown[1] that $\mathcal{Q}(\theta\,; \theta^{(i)}) \geqslant \mathcal{Q}(\theta\,; \theta^{(i-1)})$ implies $\mathcal{L}(\mathbf{X}, \mathbf{Y}\,; \theta^{(i)}) \geqslant \mathcal{L}(\mathbf{X}, \mathbf{Y}\,; \theta^{(i-1)})$, with equality only if $\theta^{(i)} = \theta^{(i-1)}$, hence ensuring convergence if $\mathcal{Q}$ is continuous and if the likelihood is bounded from above. The intermediate quantity $\mathcal{Q}$ is thus a monotonically increasing function. Under mild conditions, the stationary point is a (local) maximum so that $\nabla_{\boldsymbol{\theta}} \mathcal{L} = \mathbf{0}$, where the subscript indicates what the operator is to be taken with respect to.

**Algorithm 3.1** The Expectation-Maximization Algorithm.

1: Choose an initial value $\boldsymbol{\theta}^{(0)}$.
2: **for** $i \leftarrow 1$ until convergence **do**
3: $\quad \mathcal{Q}(\boldsymbol{\theta}\,; \boldsymbol{\theta}^{(i-1)}) \leftarrow \mathbb{E}_{\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}^{(i-1)}}(\log \mathcal{L}(\mathbf{X}, \mathbf{Y}\,; \boldsymbol{\theta}) \mid \mathbf{Y}, \boldsymbol{\theta})$ ▷ The E-step.
4: $\quad \boldsymbol{\theta}^{(i)} \leftarrow \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}\,; \boldsymbol{\theta}^{(i-1)})$ ▷ The M-step.
5: **end for**

Like maximum likelihood, the algorithm does not ensure global convergence to the global maximum as it can get stuck into a local one. The convergence criterion can for instance be based upon the change in the likelihood from one iteration to another, the value of the gradient vector, or the value of $\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}\|$ for various norms $\|\cdot\|$, e.g. the supremum norm.

### 3.2.1 Smoothing and the Backward Recursion

In addition to parameter estimation, one objective of state-space modelling is to compute optimal estimates of latent states. Such estimates are obtained as a by-product of the EM algorithm. In the E-step, the expectation of the complete log-likelihood function has to be computed given the most recent parameter estimates. The likelihood distribution $p(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\theta})$ can be computed in the same manner as in maximum likelihood estimation, viz. through the filtering method described in that context.

Parameter estimation cannot be carried out after solely solving the filtering problem as the EM algorithm relies heavily upon closed-form expressions in order to perform well. A new problem arises in this context from the need to sample from the marginal distributions $\{p(x_t \mid \mathbf{y}_T)\}$, where $t = 1, \ldots, T$, or the joint distribution $p(\mathbf{x}_T \mid \mathbf{y}_T)$, conditionally on

[1] See for instance Hamilton (1990).

the parameter estimates. The marginal distributions, which are commonly referred to as smoothing distributions, can be obtained directly from the joint distribution in question by the decomposition

$$p(\mathbf{x}_T \mid \mathbf{y}_T) = p(x_T \mid \mathbf{y}_T) \prod_{t=1}^{T-1} p(x_t \mid x_{t+1}, \mathbf{y}_T),$$

since the smoothing distributions $p(x_t \mid \mathbf{y}_T, \boldsymbol{\theta})$ can be rewritten by the law of total probability and Bayes' theorem as

$$\begin{aligned} p(x_t \mid \mathbf{y}_T, \boldsymbol{\theta}) &= \int_{\mathsf{X}} p(x_t \mid x_{t+1}, \mathbf{y}_T, \boldsymbol{\theta}) p(x_{t+1} \mid \mathbf{y}_T, \boldsymbol{\theta}) dx_{t+1} \\ &= \int_{\mathsf{X}} p(x_t \mid x_{t+1}, \mathbf{y}_t, \boldsymbol{\theta}) p(x_{t+1} \mid \mathbf{y}_T, \boldsymbol{\theta}) dx_{t+1} \\ &= \int_{\mathsf{X}} \frac{p(x_{t+1} \mid x_t, \boldsymbol{\theta}) p(x_t \mid \mathbf{y}_t, \boldsymbol{\theta}) p(x_{t+1} \mid \mathbf{y}_T, \boldsymbol{\theta})}{p(x_{t+1} \mid \mathbf{y}_t, \boldsymbol{\theta})} dx_{t+1}, \end{aligned}$$

where the second equality holds since, conditionally on $\mathbf{y}_T$, the sequence $x_t, x_{t-1}, \ldots$ is a Markov chain backward in time, see for instance Hürzeler and Künsch (2001). After identifying the distributions we immediately recognize a recursive formula, except this time the recursion is reversed and should be initialized at $t = T - 1$ and proceed until all marginal distributions have been computed. This method, which was first suggested by Kim (1994), is referred to as smoothing and is indeed a backward algorithm compared with the filter. Note that the base case where $t = T$ is obtained by filtering; consequently, marginal distributions computed by the forward algorithm has to be stored in order to carry out the smoothing.

To summarize, the E-step in the EM algorithm consists of two ingredients:

(i) a filtering procedure which computes the filtering distributions

$$p(x_t \mid \mathbf{y}_t) = \frac{p(x_t \mid \mathbf{y}_{t-1}) p(y_t \mid x_t, \mathbf{y}_{t-1})}{\int_{\mathsf{X}} p(x_t \mid \mathbf{y}_{t-1}) p(\mathbf{y}_t \mid x_t, \mathbf{y}_{t-1}) dx_t};$$

(ii) a smoothing part which, given the filter, obtains

$$p(x_t \mid \mathbf{y}_T) = \int_{\mathsf{X}} \frac{p(x_{t+1} \mid x_t) p(x_t \mid \mathbf{y}_t) p(x_{t+1} \mid \mathbf{y}_T)}{p(x_{t+1} \mid \mathbf{y}_t)} dx_{t+1},$$

where the conditional parameters have been suppressed for brevity. Estimates of the initial distribution, $\rho^{(i+1)} = \mathbb{P}(X_1 \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})$, are obtained as a by-product of the smoothing procedure. The initial distribution can also be computed directly according to equation (2.5) before executing the filter in order to speed up the convergence of the overall algorithm.

## 3.3 Monte Carlo Expectation-Maximization

When for instance closed-form expressions for some parameters are not available in a straightforward way or a smoother in the EM algorithm has not been implemented, a different approach has to be taken compared with the standard EM algorithm. New computational problems arise in this way, but they can in part be eluded by Monte Carlo (MC) methods, which is a family of methods introduced by Metropolis and Ulam (1949) that builds upon sampling and evaluation of generated samples in order to draw conclusions from a population. More formally, for an $\mathsf{X}$-valued random variable $X$ with density function $f$ the problem is to compute the expectation

$$\mathbb{E}[\phi(X)] = \int_{\mathsf{X}} \phi(x) f(x) dx,$$

which can be approximated by

$$\mathbb{E}[\phi(X)] \approx \frac{1}{N} \sum_{i=1}^{N} \phi(x_i)$$

by generating sufficiently many independent and identically distributed (i.i.d.) samples from $f$. This method is commonly referred to as basic Monte Carlo integration. By the strong law of large numbers, a sequence of Monte Carlo estimates converges $\mathbb{P}$-a.s. to the expected value as $N$ approaches infinity. Note that $\mathbb{E}[\phi(X)] = \mathbb{P}(X \in A)$ when $\phi(X) = \mathbb{1}_{\{X \in A\}}$, where $\mathbb{1}$ denotes the indicator function. The crude Monte Carlo integration has $\mathcal{O}(1/\sqrt{n})$ rate of convergence, so in order to cut the approximation error in half, the sample size has to be enlarged by a factor of four.

The slight modification of the EM algorithm using MC methods is referred to as Monte Carlo expectation-maximization (MCEM). Monte Carlo samples are usually generated through applicable Markov chain Monte Carlo (MCMC) methods such as Gibbs or Metropolis-Hastings samplers, see Algorithm 3.2 for details regarding the Gibbs sampler. Note that Gibbs sampling requires all other updated samples to be known apart from the one that should be generated.

In the case of simulating state variables by single site sampling, the conditional probabilities of the form $p(x_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{T\setminus t}, \mathbf{y}_T, \boldsymbol{\theta})$, where $\mathbf{x}_{T\setminus t} = (x_{t+1}, \ldots, x_T)$, can be rewritten as the Markov property yields

$$p(x_t \mid \mathbf{x}_{t-1}, \mathbf{x}_{T\setminus t}, \mathbf{y}_T, \boldsymbol{\theta}) = p(x_t \mid x_{t-1}, x_{t+1}, y_t, \boldsymbol{\theta}).$$

The main idea here is to simulate an entire realization $x_1, \ldots, x_T$ using the Gibbs sampler described above, conditionally on parameters and $\mathbf{y}_T$. Given a realization of the Markov chain, it is then possible to iterate new parameters, which will hopefully converge to the true ones over time by repeating the simulation of the Markov chain and parameters sequentially. This data augmentation technique, which treats the latent states as auxiliary variables missing from the observed data, is a way to solve the inference problem by augmenting data. Even though a two-stage Gibbs sampler that works on both the

parameters and the state variables is rarely required in the context of MCEM, the principle is the same regardless of whether the parameters are computed in the M-step or drawn from a certain distribution.

**Algorithm 3.2** The Gibbs Sampler.

1: Choose an initial value $x^{(0)}$.
2: **for** $i \leftarrow 1$ to $N$ **do**
3: Generate $x_1^{(i)} \sim f(x_1 \mid x_2^{(i-1)}, x_3^{(i-1)} \ldots, x_n^{(i-1)})$.
4: Generate $x_2^{(i)} \sim f(x_2 \mid x_1^{(i)}, x_3^{(i-1)}, \ldots, x_n^{(i-1)})$.
5: $\vdots$
6: Generate $x_n^{(i)} \sim f(x_n \mid x_1^{(i)}, x_2^{(i)}, \ldots, x_{n-1}^{(i)})$.
7: **end for**

Aside from the above-mentioned methods, MCMC and sequential Monte Carlo methods are also common in state-space modelling. Bayesian inference is sometimes better suited as it allows the inclusion of prior knowledge by expressing prior distributions explicitly. The EM algorithm is, however, favourable if it is possible to retrieve closed-form solutions for most parameters, since MCMC methods tend to have slow convergence. Distributions belonging to the exponential family are for this reason of certain interest, since such distributions — as we will see — ease the computational burden in the E-step.

## 3.4 Parameter Estimation via the EM Algorithm

Implementation of the iterative expectation-maximization algorithm is more cumbersome than maximum likelihood estimation. The strength of the algorithm lies in obtaining closed-form expressions for the parameters. This approach gains from distributions belonging to the exponential family, though it is not a necessity for the algorithm to be viable. Apart from the forward-backward algorithm, parameter estimates have to be computed in a sequential fashion. In contrast to maximum likelihood, the EM algorithm does not need any parameter constraints.

Some discretization scheme has to be applied in the presence of autoregressive components to make inference feasible in practice, irrespective of estimation method. After having translated the considered models in this thesis from continuous to discrete time it is possible to obtain closed-form expressions for almost all parameters.

### 3.4.1 CKLS Dynamics

Recall the CKLS model (2.2) and that it can be discretized by the Euler scheme[2]

$$\Delta Y_n = \kappa(\vartheta - Y_n)\Delta t_n + \sigma Y_n^{\gamma} \Delta W_n$$

[2]See for instance Glasserman (2003) for a thorough treatment of the Euler scheme as well as higher-order discretization schemes.

in order to get

$$Y_{n+h} = Y_n + \kappa(\vartheta - Y_n)\Delta t_n + \sigma Y_n^{\gamma}\sqrt{\Delta t_n}\epsilon_{n+h},$$

where $\Delta$ is the forward difference operator $\Delta t_n = t_{n+h} - t_n$ and $\epsilon$ is an i.i.d. standard normal random variable. On an equidistant grid we have $t_{n+h} - t_n \equiv h$ for all $n$ and consequently

$$Y_{n+h} = \kappa\vartheta h + (1 - \kappa h)Y_n + \sigma Y_n^{\gamma}\sqrt{h}\epsilon_{n+h},$$

which coincides with an AR(1) process in the case of Vašíček dynamics with $\gamma = 0$. The expectation and variance of the discretized CKLS model are

$$\mathbb{E}(Y_n) = \kappa\vartheta h + (1 - \kappa h)Y_{n-h}$$

and

$$\mathbb{V}(Y_n) = \sigma^2 Y_{n-h}^{2\gamma} h.$$

Even though an arbitrary grid spacing could be chosen, let us consider a fixed step size $h \equiv 1$ for simplicity. Note that, for applications with discrete data, the step size has to be taken sufficiently small to avoid discretization bias or at least make it arbitrarily small.

For the rest of this section I present formulae for almost all parameters by starting with CKLS parameters and later proceeding with both spike parameters and transition probabilities.

**Proposition 3.1** *Let $\{(X_t, Y_t)\}_{t=1}^T$ be an MRS model where $X$ takes values in a finite set $\mathsf{X}$. Suppose that the observation equation of regime $k \in \mathsf{X}$ has normally distributed, Euler-discretized CKLS dynamics. The model parameters obtained by the EM algorithm are then given by*

$$\kappa^{(i+1)} = \frac{\sum_{t=2}^T \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})(y_t - y_{t-1})(\vartheta^{(i)} - y_{t-1})y_{t-1}^{-2\gamma^{(i)}}}{\sum_{t=2}^T \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})(\vartheta^{(i)} - y_{t-1})^2 y_{t-1}^{-2\gamma^{(i)}}},$$

$$\vartheta^{(i+1)} = \frac{\sum_{t=2}^T \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})(y_t + (\kappa^{(i+1)} - 1)y_{t-1})y_{t-1}^{-2\gamma^{(i)}}}{\sum_{t=2}^T \kappa^{(i+1)}\mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})y_{t-1}^{-2\gamma^{(i)}}}$$

*and*

$$(\sigma^{(i+1)})^2 = \frac{\sum_{t=2}^T \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})\left[y_t - y_{t-1} - \kappa^{(i+1)}(\vartheta^{(i+1)} - y_{t-1})\right]^2 y_{t-1}^{-2\gamma^{(i)}}}{\sum_{t=2}^T \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})}.$$

*Proof.* Differentiating the expectation of the log-likelihood

$$\log p(\mathbf{X}_T, \mathbf{Y}_T \mid \boldsymbol{\theta}) = \log p(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta}) + \log p(\mathbf{X}_T \mid \boldsymbol{\theta}), \tag{3.5}$$

where the first term is

$$\begin{aligned}\log p(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta}) &= \log\left(\prod_{t=2}^{T} p(Y_t \mid X_t, \mathbf{Y}_{t-1}, \boldsymbol{\theta})\right) \\ &= \log\left(\prod_{t=2}^{T}\prod_{j\in\mathsf{X}} p(Y_t \mid X_t = j, \mathbf{Y}_{t-1}, \boldsymbol{\theta})^{\mathbb{1}_{\{X_t=j\}}}\right) \\ &= \sum_{t=2}^{T}\sum_{j\in\mathsf{X}} \mathbb{1}_{\{X_t=j\}} \log p(Y_t \mid X_t = j, \mathbf{Y}_{t-1}, \boldsymbol{\theta}),\end{aligned}$$

with respect to $\kappa$ yields

$$\left.\frac{\partial \mathcal{Q}}{\partial \kappa}\right|_{\kappa=\kappa^{(i+1)}} = \sum_{t=2}^{T} p(X_t = j \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) \frac{(y_t - y_{t-1})(\vartheta^{(i)} - y_{t-1}) - \kappa^{(i+1)}(\vartheta^{(i)} - y_{t-1})^2}{(\sigma^{(i)})^2 y_{t-1}^{2\gamma^{(i)}}}. \tag{3.6}$$

Setting equation (3.6) to zero and solving for $\kappa$ finally gives the desired expression. Formulae for the remaining parameters are obtained along the same lines. □

Contrary to the obtained formulae for the model parameters it is not possible to obtain closed-form expressions for $\gamma$. After some calculations we arrive at

$$\sum_{t=2}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) \log(y_{t-1}) \left( \frac{\left[y_t - y_{t-1} - \kappa^{(i+1)}(\vartheta^{(i+1)} - y_{t-1})\right]^2}{(\sigma^{(i+1)})^2 y_{t-1}^{2\gamma^{(i+1)}}} - 1 \right) = 0.$$

Consequently, numerical methods have to be used to estimate $\gamma$ given the remaining updated parameters.

### 3.4.2 Spike Distributions

Estimation of spike parameters is less cumbersome since spikes are assumed to be i.i.d. This section treats three distributions, namely the Gaussian, log-normal, and gamma distributions. Closed-form expressions for both Gaussian and log-normal spike parameters can be obtained readily along the same lines as in Proposition 3.1. The only difference between the two distributions is a logarithmic transformation. The perspicacious reader realizes that formulae for Gaussian spike parameters coincide with those in Proposition 3.1 in the case of Vašíček dynamics and $\kappa = 1$.

Gamma spike parameters are on the other hand a bit trickier. Since the probability density function of a random variable belonging to a gamma distribution has the form[3]

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \tag{3.7}$$

[3]N.B. that this form implies that $\alpha > 0$ and $\beta > 0$ are shape and rate parameters, respectively.

where $\Gamma$ denotes the gamma function, it is not possible to derive closed-form expressions for both variables. The rate parameter is in this case given by

$$\beta^{(i+1)} = \alpha^{(i)} \frac{\sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})}{\sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) y_t}. \tag{3.8}$$

In the derivation of the shape parameter we start by plugging equation (3.8) into equation (3.7) and then solve $\nabla_\alpha \mathcal{Q} = 0$ in order to get

$$\begin{aligned}\psi(\alpha) - \log \alpha = \sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) \left( 1 + \log y_t + \log \frac{\sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})}{\sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) y_t} \right. \\ \left. - y_t \frac{\sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})}{\sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) y_t} \right) \times \left( \sum_{t=1}^{T} \mathbb{P}(X_t = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) \right)^{-1},\end{aligned}$$

where $\psi$ denotes the digamma function defined by

$$\psi(\alpha) = \frac{d}{d\alpha} \log \Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.$$

Numerical optimization of $\alpha$ is needed since no closed-form solution exists.

### 3.4.3 Transition Probabilities

We now turn to estimation of transition probabilities. For this purpose we define $n_{ij} = \#\{1 < t \leqslant T \colon X_{t-1} = i, X_t = j\}$ as the number of transitions from state $i$ to state $j$. Due to the separation of the log-likelihood, see equation (3.5), parameter estimation of the transition probabilities does not interfere with state parameters. The following result has a very natural interpretation but is nonetheless of importance.

**Proposition 3.2** *Let $\{X_t\}_{t=1}^{T}$ be a time-homogeneous Markov chain with finite state space $\mathsf{X}$. Suppose that, for any state $i \in \mathsf{X}$, the transition probabilities $p_{ij}$ are strictly positive for all states $j$ in an indexed set $I \subset \mathsf{X}$ and zero for all states $j \notin I$. The transition probabilities obtained by the MCEM algorithm are then given by*

$$p_{ij} = \frac{n_{ij}}{\sum_{k \in I} n_{ik}}.$$

*Proof.* Consider for the moment a state space with only two states, say $i$ and $j$. Note that only the second term in equation (3.5) depends on $p_{ij}$. Rewrite this term as

$$\begin{aligned}\log p(\mathbf{X}_T \mid \boldsymbol{\theta}) &= \log \left( p(X_1 \mid \boldsymbol{\theta}) \prod_{i \in \mathsf{X}} \prod_{j \in \mathsf{X}} p_{ij}^{n_{ij}} \right) \\ &\approx \log \left( \prod_{i \in \mathsf{X}} \prod_{j \in \mathsf{X}} p_{ij}^{n_{ij}} \right) \\ &= \sum_{i \in \mathsf{X}} \sum_{j \in \mathsf{X}} n_{ij} \log p_{ij},\end{aligned}$$

where the approximation is negligible for sufficiently many observations $T$, then differentiate the expectation of the log-likelihood with respect to $p_{ii} = 1 - p_{ij}$ and solve $\nabla_{\boldsymbol{\theta}} \mathcal{Q} = \mathbf{0}$ in order to arrive at

$$\mathbb{E}(n_{ii}) \frac{1}{p_{ii}} - \mathbb{E}(n_{ij}) \frac{1}{1 - p_{ii}} = 0,$$

which is equivalent to

$$p_{ii} = \frac{\mathbb{E}(n_{ii})}{\mathbb{E}(n_{ii}) + \mathbb{E}(n_{ij})},$$

where the expectation can be taken in the Monte Carlo sense.

For a state space of higher order, a system of equations has to be solved. Expressions for transition probabilities in the case of three states are obtained straightforwardly in a similar manner. The full proof is outside the scope of this thesis and thus omitted. □

Implementation of a smoother can circumvent the Monte Carlo part in the EM algorithm. Following this approach will only lead to different calculations for the transition probabilities so that the following result has to be used in place of Proposition 3.2, while expressions for all CKLS model parameters remain the same.

**Proposition 3.3** *Let $\{(X_t, Y_t)\}_{t=1}^T$ be a time-homogeneous MRS model where $X$ takes values in a finite set $\mathsf{X}$. Suppose that, for any state $k \in \mathsf{X}$, the transition probabilities $p_{kl}$ are strictly positive for all states $l$ in an indexed set $I \subset \mathsf{X}$. The transition probabilities obtained by the EM algorithm are then given by*

$$\begin{aligned} p_{kl}^{(i+1)} &= \frac{\sum_{t=2}^T \mathbb{P}(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})}{\sum_{t=2}^T \mathbb{P}(X_{t-1} = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})} \\ &= \frac{\sum_{t=2}^T \mathbb{P}(X_t = l \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)}) p_{kl}^{(i)} \mathbb{P}(X_{t-1} = k \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}^{(i)}) \mathbb{P}(X_t = l \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}^{(i)})^{-1}}{\sum_{t=2}^T \mathbb{P}(X_{t-1} = k \mid \mathbf{y}_T, \boldsymbol{\theta}^{(i)})}. \end{aligned}$$

*Proof.* The first equality is shown by Hamilton (1990), while the second equality holds by observing that

$$\mathbb{P}(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \boldsymbol{\theta}) = \mathbb{P}(X_t = l \mid \mathbf{y}_T, \boldsymbol{\theta}) \mathbb{P}(X_{t-1} = k \mid X_t = l, \mathbf{y}_{t-1}, \boldsymbol{\theta}),$$

whence the desired result is obtained readily by applying Bayes' theorem on $\mathbb{P}(X_{t-1} = k \mid X_t = l, \mathbf{y}_{t-1}, \boldsymbol{\theta})$. □

While Proposition 3.3 is enough for time-invariant cases, some changes have to be done in order to allow for more general transitions. By contrast, the non-linear filter approach with maximum likelihood estimation taken by e.g. Filardo (1994) is more straightforward. In fact, the result below, which has some similarities to the time-invariant case, shows how parameters of time-varying transitions are to be properly estimated when the transition probabilities take the form of a multinomial logistic function. The following result is a generalization of Diebold et al. (1994), who only treat two states and logistic functions of the form $p(\mathbf{z}, \boldsymbol{\beta}) = (1 + e^{-\boldsymbol{\beta} \cdot \mathbf{z}})^{-1}$.

**Proposition 3.4** *Let $\{(X_t, Y_t)\}_{t=1}^T$ be an inhomogeneous MRS model where $X$ takes values in a finite set $\mathsf{X}$. Furthermore, let $\boldsymbol{\beta}$ be an $N$-dimensional parameter vector and let $\mathbf{z}$ be a vector of $N$ observations. Suppose that, for any state $k \in \mathsf{X}$, the transition probabilities $p_{kl}(\mathbf{z}_{t-1}, \boldsymbol{\beta}, m) = \mathbb{P}(X_t = l \mid X_{t-1} = k, \mathbf{z}_{t-1}, \boldsymbol{\beta}, m)$ are of the form*

$$p_{kl}(\mathbf{z}_{t-1}, \boldsymbol{\beta}, m) = \frac{e^{\boldsymbol{\beta}_m \cdot \mathbf{z}_{m,t-1}}}{1 + \sum_{n=1}^N e^{\beta_n z_{n,t-1}}}$$

*and that they are strictly positive for at least two states $l \in \mathsf{X}$ and zero for all other states. The parameter vector obtained by the EM algorithm is then given by*

$$\begin{aligned}\boldsymbol{\beta}_m^{(i+1)} = &\left(\sum_{t=2}^T \mathbf{z}_{m,t-1} \mathbb{P}_{\boldsymbol{\theta}^{(i)}}(X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T) \frac{\partial p_{kl}(\mathbf{z}_{t-1}, \boldsymbol{\beta}, m)}{\partial \boldsymbol{\beta}_m}\right)^{-1} \\ &\times \left(\sum_{t=2}^T \mathbf{z}_{m,t-1} \begin{Bmatrix} \mathbb{P}_{\boldsymbol{\theta}^{(i)}}(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T) - \mathbb{P}_{\boldsymbol{\theta}^{(i)}}(X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T) \\ \times \left[p_{kl}(\mathbf{z}_{t-1}, \boldsymbol{\beta}, m) - \frac{\partial p_{kl}(\mathbf{z}_{t-1}, \boldsymbol{\beta}^{(i)}, m)}{\partial \boldsymbol{\beta}_m} \boldsymbol{\beta}_m^{(i)}\right] \end{Bmatrix}\right),\end{aligned}$$

*where the vectors of partials are to be evaluated at $\boldsymbol{\beta}^{(i)}$ and given by*

$$\left.\frac{\partial p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}, m)}{\partial \boldsymbol{\beta}_m}\right|_{\boldsymbol{\beta}_m = \boldsymbol{\beta}_m^{(i)}} = \Big(p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, m, 1), p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, m, 2), \ldots, p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, m, N)\Big)$$

*and the individual partials are, in turn, given by*

$$p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, m, n) = z_{n,t-1}\left(p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, m) - p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, m)^2\right).$$

*Proof.* In contrast to time-invariant transitions as in Proposition 3.3, the problem is to solve a system of non-linear equations. The crucial part is to apply a Taylor series expansion, though the main idea is to proceed as before. Observe that the part that depends on the transitions in the E-step is

$$\mathbb{E}\left[\log p(\mathbf{X}_T \mid \boldsymbol{\theta}^{(i)})\right] = \sum_{t=2}^T \sum_{k \in \mathsf{X}} \sum_{l \in \mathsf{X}} p(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}) \log p_{kl}(t-1) \quad (3.9)$$

where exogenous dependence of $p_{kl}$ has been supressed for brevity. Note that

$$\left.\frac{\partial}{\partial \boldsymbol{\beta}_n} \log p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}, m)\right|_{\boldsymbol{\beta}_n = \boldsymbol{\beta}_n^{(i)}} = -\mathbf{z}_{n,t} p_{kl}(\mathbf{z}_t, \boldsymbol{\beta}^{(i)}, n) + \begin{cases} \mathbf{z}_{n,t} & \text{if } n = m \\ 0 & \text{if } n \neq m, \end{cases}$$

and by the law of total probability we have that

$$\sum_{l \in \mathsf{X}} p(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}) = p(X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}),$$

so differentiating the conditional expectation (3.9) with respect to $\beta_m$ yields the non-linear equations

$$\sum_{t=2}^{T} \mathbf{z}_{m,t-1} \Big\{ p(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}) \\ - p_{kl}(\mathbf{z}_{t-1}, \boldsymbol{\beta}^{(i)}, m) p(X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}) \Big\} = \mathbf{0}. \tag{3.10}$$

By approximating $p_{kl}$ by the first-order Taylor series expansion

$$p_{kl}(\boldsymbol{z}_t, \boldsymbol{\beta}^{(i)}, m) \approx p_{kl}(\boldsymbol{z}_t, \boldsymbol{\beta}^{(i)}, m) + \frac{\partial p_{kl}(\boldsymbol{z}_t, \boldsymbol{\beta}, m)}{\partial \boldsymbol{\beta}_m}\Big|_{\boldsymbol{\beta}_m = \boldsymbol{\beta}_m^{(i)}} (\boldsymbol{\beta}_m - \boldsymbol{\beta}_m^{(i)})$$

and plugging this into the non-linear equations (3.10) we obtain the linear equations

$$\sum_{t=2}^{T} \mathbf{z}_{m,t-1} \Bigg\{ p(X_t = l, X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}) - p(X_{t-1} = k \mid \mathbf{y}_T, \mathbf{z}_T, \boldsymbol{\theta}^{(i)}) \\ \times \left[ p_{kl}(\boldsymbol{z}_{t-1}, \boldsymbol{\beta}^{(i)}, m) + \frac{\partial p_{kl}(\boldsymbol{z}_{t-1}, \boldsymbol{\beta}, m)}{\partial \boldsymbol{\beta}_m}\Big|_{\boldsymbol{\beta}_m = \boldsymbol{\beta}_m^{(i)}} (\boldsymbol{\beta}_m - \boldsymbol{\beta}_m^{(i)}) \right] \Bigg\} = \mathbf{0},$$

which can be solved for $\boldsymbol{\beta}_m$. □

**Remark 3.5** *The joint smoothing distributions* $\mathbb{P}(X_t = j, X_{t-1} = i \mid \mathbf{y}_N, \mathbf{z}_N, \boldsymbol{\theta})$ *in Proposition 3.4 are given by Proposition 3.3.*

In order to improve the convergence of the parameter estimates using the EM algorithm, choose the initial logistic coefficients according to the following scheme. First, run the EM algorithm for the case of a time-homogeneous MRS model to obtain $\hat{p}$. Second, find the inverse of $\hat{p}$, i.e. calculate the corresponding explanatory parameter for the intercept as the logit function

$$\beta^{(0)} = \log(\hat{p}) - \log(1 - \hat{p})$$

while neglecting any other exogenous data. Finally, extend the logistic function to a preferable number of variables, either with one or several variables at a time.

## 3.5 Method Validation via Simulations

Both maximum likelihood and expectation-maximization are validated and compared to each other by numerous simulations, of which some shall be presented to set an example of the simulation results.

### 3.5.1 An Inhomogeneous Two-State Model

The first example is an inhomogeneous two-state MRS model with CKLS dynamics with parameters $\kappa = 0.3$, $\vartheta = 2.5$, $\sigma = 0.15$ and $\gamma = 0.5$ in the normally distributed base regime, and Gaussian spikes with parameters $\mu_{_S} = 4.2$ and $\sigma_{_S} = 1.2$. I let the transition probability matrix be

$$\mathbf{P} = \begin{bmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{bmatrix},$$

where

$$p_i = \left[1 + \exp(-\beta_{_{p_i,0}} - \beta_{_{p_i,1}} z)\right]^{-1} \tag{3.11}$$

and the coefficients are $\beta_{_{p1,0}} = 4$, $\beta_{_{p1,1}} = 2$, $\beta_{_{p2,0}} = 3$ and $\beta_{_{p2,1}} = 1.5$. In addition, the exogenous input $z$ is taken as a smooth sine function, $z(t) = \sin(2\pi t/250)$, to simulate seasonal variations.

As Frühwirth-Schnatter (2006) points out in her monograph, maximum likelihood estimation suffers some difficulties when it comes to finding global maxima, particularly for small sample sizes. Loosely speaking, estimation problems obviously grow with the number of parameters and some parameters are harder to estimate than others. What sample size suffices to obtain good estimates is hard to say as it depends on the nature of the parameters. Needless to say, in order to estimate transition probabilities, at least a few transitions are required to have occurred. With this in mind, a relatively large sample — $T = 10{,}000$, which is in parity with usual simulation studies — is chosen to facilitate estimation of transition probabilities, since sufficiently many transitions have to occur. The sample size shall in particular improve estimation of logistic coefficients, as these are obviously harder to estimate than constant transition probabilities.

Most parameter estimates are obtained by Propositions 3.1 and 3.4 when using the EM algorithm, which is known to obtain rough approximations in relatively few iterations, but convergence is on the other hand known to be slow. For this reason the algorithm is terminated after one hundred iterations, since this should yield relatively good approximations. The algorithm actually satisfies some robust convergence criteria in less iterations.

A very clear and typical convergence of parameter estimates is illustrated in Figure 3.1. The rate of convergence of some parameters, especially the logistic coefficients, is significantly slower. The slow convergence of these coefficients is probably due to the approximation by Taylor series, see Proposition 3.4. In this case the plots show that the EM algorithm has a higher rate of convergence than maximum likelihood. Bear in mind that some estimates are more sensitive to initial parameter values than others; even small perturbations of initial parameter values can result in slower convergence or larger estimation errors.

The convergence of parameters does not reveal the magnitude of the estimation errors. For this reason the box plot that is shown in the left panel in Figure 3.2 clearly illustrates that both methods perform very well with relatively small estimation errors.

The estimation errors are in this case $|\hat{\boldsymbol{\theta}}| - |\boldsymbol{\theta}_\star|$, i.e. the difference between the parameter estimates and the true parameters.

Finally, log-likelihood functions obtained by both methods are plotted for completeness, see Figure 3.3. The likelihood functions are clearly monotone functions and both ML and EM reach their maxima relatively quickly.

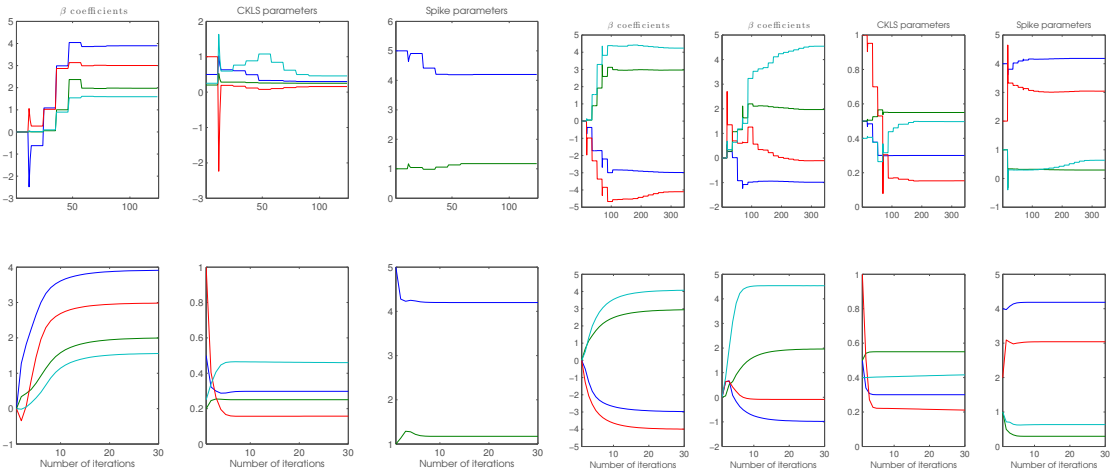



Figure 3.1: Convergence of parameters using ML (upper row) and the EM algorithm (lower row). The first three columns show parameters of an inhomogeneous two-state MRS model, of which the $\beta$ coefficients of the transition probabilities are displayed in the first column, whereas CKLS and spike parameters are shown in the second and third columns, respectively. Trajectories of parameter estimates of an inhomogeneous three-state MRS model are depicted in the last four columns, where the fifth and sixth columns display $\beta$ coefficients of the transition probabilities from the base and spike regimes, respectively. Moreover, the seventh and eighth columns show CKLS and spike parameters, respectively. In order to easier fit the plots, $\kappa$ has been scaled with one and two orders of a magnitude in the case of two and three states, respectively.

### 3.5.2 An Inhomogeneous Three-State Model

An inhomogeneous three-state MRS model with similar CKLS parameters as in the case of two states, except that $\vartheta = 55$, shall epitomize the performance of the estimation methods. The spikes are also changed to be log-normal and $\sigma_S$ is substituted with $\sigma_S = 0.3$. The extended model not only introduces log-normal drops with the parameters $\mu_D = 3$ and $\sigma_D = 0.6$, but it also extends the transition probability matrix to

$$\mathbf{P} = \begin{bmatrix} 1 - p_1 - p_2 & p_1 & p_2 \\ p_3 & 1 - p_3 & 0 \\ p_4 & 0 & 1 - p_4 \end{bmatrix},$$

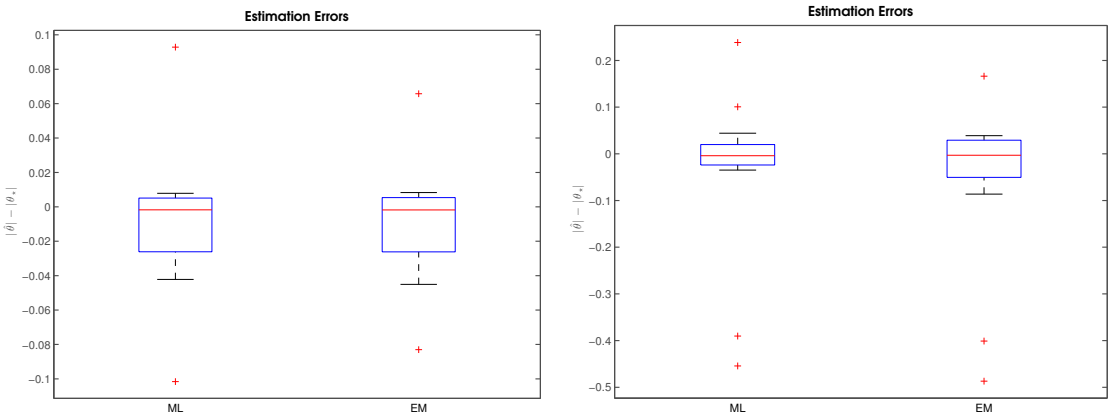



Figure 3.2: Box plots for differences between final parameter estimates and true parameters, both in absolute terms, as obtained by the MLE and the EM algorithm. The plots illustrate that the errors for an inhomogeneous two-state (left panel) as well as three-state (right panel) MRS model are approximately of the same size, irrespective of estimation method.

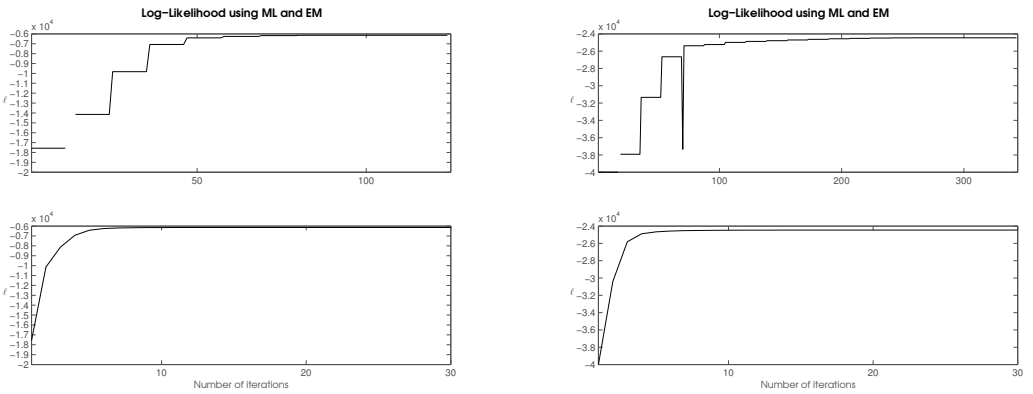



Figure 3.3: Convergence of log-likelihood functions obtained by ML and the EM algorithm showcased in the left and right plots for an inhomogeneous two-state and three-state MRS model, respectively.

where

$$p_i = \frac{\exp(\beta_{p_i,0} + \beta_{p_i,1} z)}{1 + \sum_{i=1}^{2} \exp(\beta_{p_i,0} + \beta_{p_i,1} z)}$$

for $i = 1, 2$ and $p_i$ is of the form (3.11) for $i = 3, 4$. The rationale behind this specification, which does not allow one-step transitions between the spike and drop states, is its suitability for modelling electricity prices. The $\beta$ coefficients of the transition probabilities are chosen as $\beta_{p_1,0} = -3$, $\beta_{p_1,1} = 3$, $\beta_{p_2,0} = -4$, $\beta_{p_2,1} = 4$, $\beta_{p_3,0} = -1$, $\beta_{p_3,1} = 2$, $\beta_{p_4,0} = -0.5$ and $\beta_{p_4,1} = 5$. The exogenous process is once again chosen as the sinusoid specified above.

The introduction of a new state increases the computational burden and results in

slightly slower convergence of parameter estimates, see Figure 3.1. While the convergence of the log-likelihood functions is still relatively fast for the EM algorithm, it is clearly slower for maximum likelihood, see Figure 3.3. The parameter estimation also suffers from the six new parameters, see Figure 3.2, but still achieves satisfactory estimates nevertheless.

The simulation results conclude that both estimation methods perform well and are able to achieve global convergence. Although the extension to three states complicates parameter estimation, both methods cope with the computational burden and deliver relatively fast convergence, especially the EM algorithm. Needless to say, estimation errors are naturally larger for three-state models given the same number of iterations, but simulations show that the estimation errors are overall relatively small, hence making inference in both MRS models promising.

## 3.6 Model Selection

Determining the optimal model order is a notoriously difficult task. There are numerous estimators and techniques to use and while no estimator is undoubtedly the best, some are more appropriate to use than others depending on the problem. Three robust tests which are useful for determining the model order complete this chapter.

### 3.6.1 Likelihood Ratio Test

Comparison of how well two models fit data and evaluation of parameter significance are two frequently used ingredients in order to draw conclusion from the models. A very commonly used method in this context is the likelihood ratio (LR) test, which is an asymptotic test based upon two different models. The parameter space of the so-called restricted model is a subspace of the parameter space of another model called the unrestricted model. Suppose that we would like to test the hypotheses

$$\begin{aligned} H_0 &: \theta \in \Theta_0, \\ H_1 &: \theta \notin \Theta_0, \end{aligned}$$

for an estimate of $\theta$ and parameter space $\Theta_0 \subset \Theta$. A simple way to do this is to test the null hypothesis against the alternative hypothesis by defining the likelihood ratio

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\mathbf{y}\,;\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\mathbf{y}\,;\theta)}.$$

By the definition the denominator will be dominating if the null hypothesis does not hold. The LR test statistic is defined as twice the logarithm of $\Lambda$ so it holds asymptotically that

$$-2\log\Lambda = -2\left(\sup_{\theta \in \Theta_0} \ell(\mathbf{y}\,;\theta) - \sup_{\theta \in \Theta} \ell(\mathbf{y}\,;\theta)\right) \sim \chi^2(\nu),$$

where $\nu$ is the difference between the number of free parameters of the restricted and unrestricted models. In applications, hypothesis testing is often conducted in the case of $\Theta_0$ being a singleton.

### 3.6.2 Generalized Information Criterion

Besides the likelihood ratio test mentioned above, other statistics and methods may be of interest. Information theoretic models also emphasize the determination of optimal model order. The rationale behind such models is to favour parsimony by minimizing a penalized likelihood function $\mathcal{L}(\mathbf{y}_T\,;\hat{\boldsymbol{\theta}}_p)$, where $p$ denotes the number of model parameters, given sufficiently many observations $T$. The optimization problem is to be taken with respect to the number of parameters, which means that the penalized likelihood function is a function of this number as well. The generalized information criterion (GIC), which is defined by

$$\mathrm{GIC}(p) = \alpha_{p,T} - 2\ell(\mathbf{y}_T\,;\hat{\boldsymbol{\theta}}_p), \tag{3.12}$$

unifies the idea of such penalized likelihood functions in a compact way. The optimization problem yielding the optimal model order thus becomes

$$\hat{p} = \underset{\{p\}}{\arg\min}\,\mathrm{GIC}(p).$$

Equation (3.12) coincides with the Akaike information criterion (AIC) for $\alpha_{p,T} = 2p$ and the Kullback information criterion (KIC) for $\alpha_{p,T} = 3p$. Moreover, when $\alpha_{p,T} = p\log T$ it is referred to as the Bayesian information criterion (BIC) (or the Schwarz criterion after the work of Schwarz (1978)), which provides a strongly consistent Markov order estimator as shown by Csiszár and Shields (2000).

### 3.6.3 Kolmogorov-Smirnov Test

The goodness of fit can be tested with the non-parametric Kolmogorov-Smirnov (K-S) test. The one-sample Kolmogorov-Smirnov test statistic is defined by

$$K = \sup_x |F_n(x) - F(x)|,$$

where $F_n$ is the empirical cumulative distribution function (cdf) and $F$ is the theoretical cdf. From the definition it is clear that the test statistic measures the supremum distance between two cdfs. It can also be used for two or more samples by replacing the theoretical cdf with the cdf of the second sample. The null hypothesis assumes that samples are drawn from the same distribution. The Lilliefors test, which is a special case of the K-S test, is a normality test and can for example be used to test the assumption that innovations are standard normal.

Chapter 4

# Modelling Electricity Prices

## 4.1 Model Framework

A popular way of describing the price dynamics in electricity markets is to use Markov-switching models. Most applications that model electricity prices assume at most a handful of states. Two or three states are usually enough to incorporate the desired dynamics of the models, see e.g. Ethier and Mount (1998), Janczura and Weron (2010, 2012), de Jong (2006), de Jong and Huisman (2002), de Jong and Schneider (2009), Lindström and Regland (2012), and Weron et al. (2004). States are quite often representing economic conditions such as stability and crisis, thereby the small number of states. Lack of sufficiently many transitions between potential states may also be a reason for not investigating larger numbers of states, since estimations would not benefit from superfluous regimes. As mentioned earlier, at least a few transitions have to take place in order to get proper estimates of transition probabilities and, in turn, any explanatory variables.

With all this in mind, electricity prices are modelled in a similar fashion. Formally, we let $\{(X_t, Y_t)\}_{t=1}^T$ be an MRS model where the real-valued price process $\{Y_t\}_{t=1}^T$ is driven by a latent process $\{X_t\}_{t=1}^T$ that takes values in a finite set $\mathsf{X}$. In order to avoid equivocation, the underlying Markov chain is assumed to be ergodic so that the long-term evolution of the chain is independent of its initial distribution. The state space in which the Markov chain takes it values represents the different price dynamics, for instance mean-reverting and extreme prices. In addition to a base regime (B) representing stable prices, extreme price movements are described by spikes (S) and drops (D) such that $\mathsf{X} = \{B, S, D\}$. There is unfortunately no consensus of what the definition of a spike (or drop for that matter) really is. In reality there exist different definitions, see for instance Weron (2006) and Janczura et al. (2013). One advantage of MRS models is that spikes do not have to be predefined for the model calibration. In this thesis I define spikes and drops as identified by an MRS model, hence different models will detect different spikes.

## 4.2 First Generation Models

An extensive collection of price models is presented in Janczura and Weron (2010), wherein models are divided into two distinct categories, viz. first and second generation models. The fundamental differences between the two categories are statistical innovations and exogenous input; both shall be treated exhaustively in the coming sections. First generation models share many issues with second generation models, hence a proper discussion of the modelling approaches is in order.

The first studies that examined electricity prices in a regime-switching framework were Deng (1998) and Ethier and Mount (1998). The predecessors usually assumed two states so that the transition probability matrix takes the form

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}.$$

Like in the case of Hamilton (1989), this specification allows for a base regime and a spike regime with different dynamics.

Transitions from the spike to the base regime usually take two forms; either they are exponentially decaying or they are abrupt. Huisman and Mahieu (2003) model transitions from a spike state to a base state in a novel way by letting an intermediate regime control the price movement back to the base state. Their three-state model, however, does not allow consecutive spikes and is limited to Gaussian spikes, while Poisson jumps are only considered for jump-diffusions. Weron et al. (2004) further modify the model by specifying a two-state model with log-normal spikes.

A model framework that has received considerable attention in the literature consists of independent regime models. The idea of those models is, as the name suggests, to let all regime dynamics be mutually independent of each other, thereby superseding the model of Huisman and Mahieu (2003). A problem that arises from this model formulation is to determine the penultimate value $y_{k,t-1}$ of an output processes with autoregressive components in state $k$, conditionally on the observed values $\mathbf{y}_t$. Due to the importance and popularity of this setting, an entire section is dedicated wholeheartedly to independent regime models.

Base regime dynamics are known to be normally distributed and, as a matter of fact, most studies in the literature work under the normality assumption. Moreover, as Eydeland and Wolyniec (2003) among others point out, spot prices are non-stationary and some sort of deseasonalization is necessary in order to base analysis upon stationary price processes. As Janczura et al. (2013) emphasize, this is far from a trivial task and probably one of the greatest issues in modelling electricity prices due to the broad spectrum of seasonal frequencies of spot prices. For this reason an entire section is devoted to trends and seasonality.

## 4.3 Price Transformations

A topic discussed immensely by the literature is what kinds of transformations are needed in order to obtain time series that are easy to work with. The issue is complicated further by the fact that new regulations on wholesale trading has enabled negative prices — spot prices in EEX have since 2008 been allowed to be negative. Negative prices are a combination of low demand and high supply, which can occur due to significant boosts in wind and solar power production, limited transmission and storage capacities, and by enacting new regulatory frameworks, to name a few reasons. They are rational from a short-term economic perspective, since costs of rescheduling power generation may be higher.

The consensus is to apply the natural logarithmic transformation to spot prices, $p_t$, though this approach is not necessarily the optimal, cf. Weron (2009), who suggests to model observed prices rather than log prices, thus evading the problem of logarithmically transforming negative prices. A major problem with the ubiquitous log function is that it is undefined for $p = 0$ and complex-valued for $p < 0$. Some sort of shift or truncation has to be carried out as a consequence, in order to allow for logarithmic transformation.

From the fact that negative prices have been observed it is clear that the log function has to be replaced with a different transformation such that negative prices will be real-valued and yet attain the variance stabilization provided by the log function. There exist quite many suggestions of how negative prices should be treated, such as truncating negative prices to the positive real numbers, but such approaches are not statistically sound as they are likely to distort price information. Keles et al. (2012) suggest for instance to truncate negative prices to the smallest tick value and then rescale log prices with their observed signs using a novel retransformation technique. The main drawback of this work-around is loss of spike information and its limitation to the market-specific retransformation method. In a more natural way to deal with negative prices, Schneider (2012) proposes the inverse hyperbolic sine as this odd function has a similar shape compared to the logarithmic function, but also because it does not encounter problems with negative prices.

One suggestion is to use the transformation

$$f(p) = \operatorname{sgn}(p) \log(|p| + 1), \tag{4.1}$$

which has more similarities to the log function than the inverse hyperbolic sine function used by Schneider (2012), cf. Figure 4.1. The proposed transformation (4.1) is also an odd function but contrasts to $\sinh^{-1}(p) = \log(p + \sqrt{p^2 + 1})$, where

$$\lim_{|p| \to \infty} \operatorname{sgn}(p) \frac{\operatorname{arcsinh}(p)}{\log|2p|} = 1,$$

as it asymptotically holds that $\log(|p|+1)/\log|p| \to 1$ as $|p| \to \infty$, that is the modification (4.1) converges to the popular log function. Figure 4.1 also clearly illustrates the superiority of the shifted log function (4.1) over the inverse hyperbolic sine in terms of mimicking

the log function. By applying Itô's lemma to equation (4.1) for positive prices having Vašíček dynamics we end up with

$$df(p_t)(p_t+1) = \kappa\left(\vartheta - \frac{\sigma^2}{2\kappa(p_t+1)} - f(p_t)\right)dt + \sigma dW_t,$$

which shows that the transformation only affects the drift term.

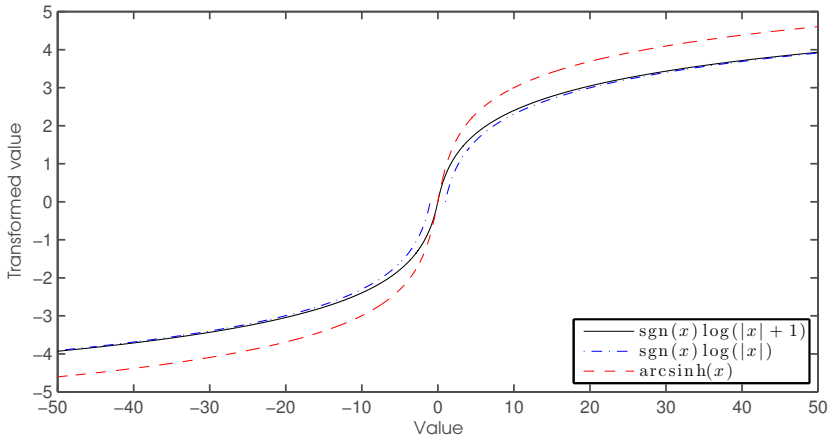



Figure 4.1: Comparison between the shifted log function (4.1) and the inverse hyperbolic sine function proposed by Schneider (2012). The former fits the standard log function much better for both positive and negative values.

For the Nordic wholesale market, which by contrast has not experienced any negative spot prices, I comply with the main opinion, since it leads to variance stabilization and has been successful in many price models. Incidentally, note that by assuming Vašíček dynamics and log prices, $y_t = \log p_t$, the dynamics coincide with the Black and Karasinski (1991) model. For EEX I use the inverse hyperbolic sine function suggested by Schneider (2012), since it has already proven useful. This thesis also utilizes the remedy used by Janczura and Weron (2012) that substitutes the diffusion term in the dynamics (2.2) with $\sigma|y_t|^\gamma dW_t$, thus avoiding negative variance that otherwise arises from negative prices.

## 4.4 Deseasonalization

Electricity prices are highly exposed to seasonal trends. A standard way of representing prices is by a deterministic seasonal component, $d_t$, and a stochastic component, $w_t$, such that the log price itself is $y_t = d_t + w_t$. We would like to filter out the seasonal component and solely model the stochastic component. There are different suggestions to achieve this goal. Janczura et al. (2013) emphasize the importance of deseasonalization and argue that this is the key to carry out a proper analysis. Lucia and Schwartz (2002) propose both dummy variables and sinusoidal functions to model seasonal patterns. Sinusoids coupled with an exponential moving average is used by Janczura et al. (2013).

An extensive comparison of just over 300 models estimating and forecasting a long-term seasonal component is conducted by Nowotarski et al. (2013). The thorough study, which comprises of a whole range of Fourier and wavelet-based decompositions, shows that wavelet-based models perform best, while the choice of wavelet family — Coiflets or Daubechies — is not critical.

A quite different approach that has proven successful is to model the mean-reversion level directly with forward prices. This approach, which was introduced by Blanco et al. (2002) and has been prosperous in e.g. de Jong and Schneider (2009) and Lindström and Regland (2012), is in principle equivalent to subtracting forward prices from the spot prices to obtain a stationary process since forward prices contain both monthly and seasonal trends. This stems from the definition of forward contracts on electricity prices which is quite different compared with equity. Under the risk-neutral probability measure $\mathbb{Q}$, the price at time $t$ of a common forward contract with time of maturity $T$ is calculated as

$$p_f(t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}(\Phi(S(T)) \mid \mathcal{F}_t),$$

where $r$ is the risk-free interest rate, $S$ is the underlying asset, and $\Phi$ is a contingent claim. The pricing convention of forwards on stocks do not apply to forward contracts on electricity spot prices as $\Phi$ has a remarkably different form. The price at time $t$ of a forward on electricity spot prices with time of maturity $T$ is the expected average of the spot prices $\{p_s(u) \colon t \leqslant u \leqslant T\}$ and is given by the risk-neutral valuation formula as

$$p_f(t) = e^{-r\tau}\mathbb{E}^{\mathbb{Q}}\left(\frac{1}{\tau}\int_t^{t+\tau} p_s(s')ds' \;\middle|\; \mathcal{F}_t\right),$$

where $\tau = T - t$ is the time to maturity. From this definition it is clear that forward contracts on electricity spot prices are path-dependent and forward-looking, hence including much of the information about future spot prices. In this thesis I let the day-ahead spot price follow the monthly forward price.

To this end daily as well as hourly trends remain. The former is filtered out after applying dummy variables with weekends corresponding to one. Parameters are then estimated by multiple linear regression and subtracted from the log price spread. Hourly trends are processed in the same manner with the slight difference that an intercept is excluded due to multicollinearity. The rationale behind this approach is consistency with earlier studies, see for instance Haldrup and Nielsen (2006), Huisman (2008), and Lucia and Schwartz (2002), which makes it easier to compare model selection rather than deseasonalization techniques.

## 4.5 Independent Regimes

Common model specifications capture price hikes by letting spikes be independent of the base regime, even though the transition from a spike state to an autoregressive state is relatively slow as such models presume an exponential growth or decay depending

on the previous state. Exponential decay is for instance evident when higher prices are followed by mean-reversion to a lower and more stable price level. This framework also naturally captures negative shocks. N.B. that no constraint is put on the number of states exhibiting lagged dependencies, hence enabling spikes to be modelled as for instance AR processes.

Both intraday and daily electricity prices have been observed to move between different price dynamics in an abrupt manner. A jump-diffusion process is not enough to capture this peculiar behaviour since it does not allow for consecutive spikes and does not capture the jumps in magnitude and time. For this reason a modification of the MRS models was proposed by de Jong and Huisman (2002), where transitions are not slowly growing or decaying to a mean-reversion level but rather immediately jumping to a new price level. This setting, which captures sharp bidirectional jumps and thereby supersedes the model of Huisman and Mahieu (2003), is referred to as independent regime (IR) models, or independent spike models, and has been successful in the modelling of electricity spot prices, see e.g. de Jong and Huisman (2002), de Jong (2006), de Jong and Schneider (2009), Janczura and Weron (2010, 2012), and Lindström and Regland (2012). For a graphical illustration of the model see Figure 4.2.

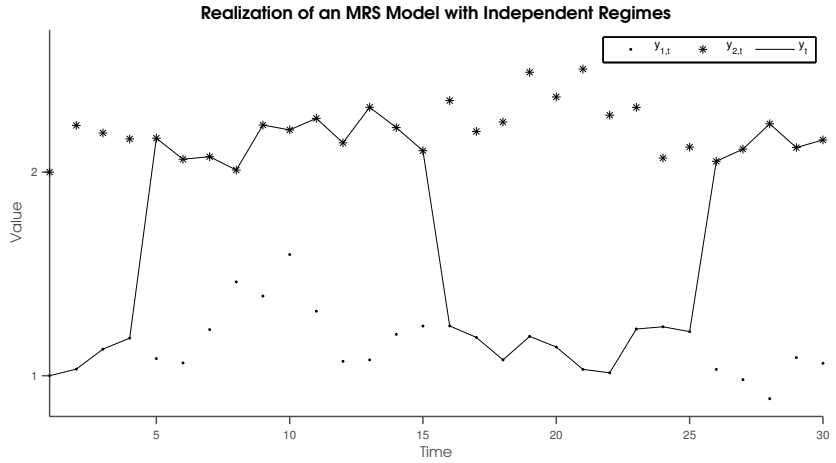



Figure 4.2: Realization of a simulated two-state Markov-switching model with independent regimes depicting abrupt jumps from one process to another.

In contrast to traditional models, a problem arises when lag-dependent processes are not observable, but since they depend on lagged values, e.g. $\mathbf{y_{k,t-1}}$ for state $k$, every such process has to be updated in the IR setting in order to make sure every lagged value comes from the corresponding process. Here the notation $\mathbf{y_{k,t}}$ denotes the trajectory of the output process in state $k$. One can think of all output processes being latent except for the currently observed one. The remaining processes are determined by estimating future trajectories from the last observed value of each process. The distribution of $Y_t$ thus depends on all its lagged values $y_1, \ldots, y_{t-1}$ and therefore all $K^t$ possible outcomes of $(x_1, \ldots, x_t)$ have to be evaluated, where $K$ is the cardinality of $\mathsf{X}$. It is impossible

to store all probabilities of future trajectories due to computational limitations. For this reason de Jong and Huisman (2002) argue that only probabilities of the last ten observations have to be stored. By contrast, Janczura and Weron (2012) suggest a more parsimonious approach which simultaneously does not presume $p_{ij}^{(n)}$ to be sufficiently small for any $n \leqslant l$, where $l$ is the maximum number of lags that has to be stored. One of the main drawbacks of the suggestion by de Jong and Huisman (2002) to use e.g. $l = 10$ as the upper boundary is highlighted by Janczura and Weron (2012). Such a restriction on the number of lags may not necessarily be enough in order to ensure that any of the $l$ lagged values $\mathbf{y}_{t-1\setminus t-l-1}$ have visited a desired state at least once. This uncertainty increases with the number of states. Instead they propose to use the expectation of the latent output variables, i.e. $\mathbb{E}(Y_{k,t-1} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta})$ for each state $k$ with autoregressors, to replace the unobservable $y_{k,t-1}$. The following result of Janczura and Weron (2012) shows how to cope with this problem.

**Proposition 4.1** *Let $\{(X_t, Y_t)\}_{t\geqslant 0}$ be an MRS model and suppose that the observation equation of regime $k$ has CKLS dynamics. Then it holds that*

$$\begin{aligned}\mathbb{E}(Y_{t,k} \mid \mathbf{y}_t, \boldsymbol{\theta}^{(i)}) =& \mathbb{P}(X_t = k \mid \mathbf{y}_t, \boldsymbol{\theta}^{(i)}) y_t + \mathbb{P}(X_t \neq k \mid \mathbf{y}_t, \boldsymbol{\theta}^{(i)}) \\ &\times \kappa^{(i)} \left\{ \vartheta^{(i)} - \mathbb{E}(Y_{t-1,k} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}^{(i)}) \right\}. \end{aligned} \tag{4.2}$$

*Proof.* Following Janczura and Weron (2012) we have that

$$Y_{t,k} = \mathbb{1}_{\{X_t = k\}} Y_t + \mathbb{1}_{\{X_t \neq k\}} \left( \kappa(\vartheta - Y_{t-1,k}) + \sigma Y_{t-1,k}^{\gamma} \epsilon_t \right),$$

for which the conditional expectation given $\mathbf{y}_t$ and $\boldsymbol{\theta}^{(i)}$ is

$$\begin{aligned}\mathbb{E}(Y_{t,k} \mid \mathbf{y}_t, \boldsymbol{\theta}^{(i)}) =& \mathbb{P}(X_t = k \mid \mathbf{y}_t, \boldsymbol{\theta}^{(i)}) Y_t + \mathbb{P}(X_t \neq k \mid \mathbf{y}_t, \boldsymbol{\theta}^{(i)}) \\ &\times \begin{Bmatrix} \kappa^{(i)}(\vartheta^{(i)} - \mathbb{E}(Y_{t-1,k} \mid \mathbf{y}_t, X_t \neq k, \boldsymbol{\theta}^{(i)}) \\ -\sigma^{(i)} \mathbb{E}(Y_{t-1,k}^{\gamma^{(i)}} \epsilon_t \mid \mathbf{y}_t, X_t \neq k, \boldsymbol{\theta}^{(i)}) \end{Bmatrix}. \end{aligned} \tag{4.3}$$

Note that

$$\mathbb{E}(Y_{t-1,k} \mid \mathbf{y}_t, X_t \neq k, \boldsymbol{\theta}^{(i)}) = \mathbb{E}(Y_{t-1,k} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}^{(i)})$$

as well as

$$\mathbb{E}(Y_{t-1,k}^{\gamma^{(i)}} \epsilon_t \mid \mathbf{y}_t, X_t \neq k, \boldsymbol{\theta}^{(i)}) = \mathbb{E}(Y_{t-1,k}^{\gamma^{(i)}} \epsilon_t \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}^{(i)}).$$

From the tower property,

$$\mathbb{E}\left[ \mathbb{E}(\xi \mid \mathcal{G}) \,\middle|\, \mathcal{H} \right] = \mathbb{E}(\xi \mid \mathcal{H}) \quad \text{if} \quad \mathcal{H} \subset \mathcal{G},$$

where $\xi$ is an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ and $\mathcal{H}$ are $\sigma$-algebras on $\Omega$ contained in $\mathcal{F}$, we have that

$$\mathbb{E}(Y_{t-1,k}^{\gamma^{(i)}} \epsilon_t \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}^{(i)}) = 0,$$

which plugged into equation (4.3) yields the desired formula. □

Implementation of the IR framework is straightforward given Proposition 4.1. The outlined routine does not change much expect for the filtering part, where now the conditional expectations have to be computed in a sequential manner according to equation (4.2). All these expectations have to be stored in order to replace lagged values in the M-step, cf. Proposition 3.1.

## 4.6 Second Generation Models

Second generation models address two main objectives; they try to extend the modelling framework by using fundamental information, and construct statistical refinements in order to improve estimations. While a discussion devoted to models with exogenous input shall be presented in the subsequent section, this section is intended for the complete model dynamics together with statistical innovations.

### 4.6.1 Model Specification

Up until now we have doggedly hold on to the CKLS model even though it does not have closed-form solutions in continuous time for all $\gamma$ and its non-linearity aggravates parameter estimation. The rationale behind this choice is the generalized form of the model and its ability to capture local volatility, which can improve the goodness of fit. Results from e.g. Janczura and Weron (2010) and Lindström and Regland (2012) show that CKLS dynamics occasionally capture the inverse leverage effect and that price sensitivity is better explained by CIR-type rather than Vašíček dynamics.

Considering that monthly forward prices are used in the deseasonalization step, a slight modification of the CKLS model is necessary in order to adhere to this approach. On this front we let the mean-reversion level depend on the forward price process $u_t$ so that $\vartheta$ is time-dependent rather than constant, i.e. $\vartheta_t \equiv \eta u_t$. Coincidentally, with this formulation the dynamics in the base regime equal the Hull and White (1990) one-factor model. Neither $\kappa$ nor $\sigma$ in Proposition 3.1 are affected more than by the aforementioned transformation. By contrast, the constant $\eta$ becomes

$$\eta^{(i+1)} = \frac{\sum_{n=2}^{N} \mathbb{P}(x_n = k \mid \mathbf{y}_N, \boldsymbol{\theta}^{(i)}) u_{t-1}(y_t + (\kappa^{(i+1)} - 1)y_{t-1}) y_{n-1}^{-2\gamma^{(i)}}}{\sum_{n=2}^{N} \kappa^{(i+1)} \mathbb{P}(x_n = k \mid \mathbf{y}_N, \boldsymbol{\theta}^{(i)}) u_{t-1}^2 y_{n-1}^{-2\gamma^{(i)}}}.$$

Like Janczura and Weron (2010) among others, we aim for three states and the shifted log-normal distribution, but in order to comply with possible non-stationarity we adopt the spike dynamics that are used by e.g. de Jong and Schneider (2009) and Lindström and Regland (2012) so that the complete price dynamics become

$$\begin{aligned} y_{B,t+1} &= \kappa\vartheta_t + (1-\kappa)y_{B,t} + \sigma|y_{B,t}|^{\gamma}\epsilon_{B,t+1}, & \epsilon_{B,t} &\sim \mathcal{N}(0,1), \\ y_{S,t+1} &= \vartheta_t + v_{S,t+1}, & v_{S,t} &\sim F_S, \\ y_{D,t+1} &= \vartheta_t - v_{D,t+1}, & v_{D,t} &\sim F_D, \end{aligned} \tag{4.4}$$

where $F$ is a generic distribution. For some distributions the shifts in the middle and lower equations of (4.4) imply that zero probability is assigned to prices below (or above) the price spread, since a random variable $\xi$ that belongs to $F$ only has support on $\xi > 0$, thus ensuring spikes (or drops) in one direction. From the complete model dynamics (4.4) it is clear that spikes and drops may belong to different families of distributions, i.e. $F_S \neq F_D$. The tradition, however, has been to assume spikes and drops belonging to the same family of distributions. This thesis continues that tradition.

Regland and Lindström (2010) show that, for some markets, gamma spikes fit data better than log-normally distributed spikes. The gamma distribution has fatter tails than the Gaussian distribution but thinner tails than the log-normal one, which is so heavy-tailed that the conditional expectation of spot prices is undefined. For this reason gamma spikes are of uttermost interest, although other distributions such as extreme value-based distributions like the Burr and Pareto distributions, the normal inverse Gaussian (see. e.g. Weron (2009)), the beta distribution (see. e.g. Becker et al. (2007)), the gamma-related Nakagami distribution, or even the Rice distribution could all be of interest as well. In fact, Janczura and Weron (2010) compare the Pareto and log-normal distributions and find that the latter performs better.

With slightly different semiparametric Markov-switching (SMS) models, Eichler and Türk (2013) do not specify observation equations of spikes, instead they try to fit data to unspecified spikes. Estimation is in this way not restricted to known distributions and does not venture individual parameters to be unidentified, which might be the case when considering for instance compound Poisson processes.

In order to make the estimation process more tractable and ease the computational burden, the model is relaxed by assuming some transition probabilities to be zero. More specifically, the three states of the MRS models are thus toggled by

$$\mathbf{P} = \begin{bmatrix} 1 - p_{BS} - p_{BD} & p_{BS} & p_{BD} \\ p_{SB} & 1 - p_{SB} & 0 \\ p_{DB} & 0 & 1 - p_{DB} \end{bmatrix},$$

such that one-step transitions from spikes to drops and vice versa are not permitted. This assumption is in line with earlier studies, see e.g. de Jong and Schneider (2009) and Lindström and Regland (2012), and seems reasonable in order to avoid perplexity. The transition probabilities can be either constant or time-varying. In the latter case they vary over time according to some multinomial logistic functions of the form

$$p_{Bi} = \frac{e^{\boldsymbol{\beta}_{Bi} \cdot \mathbf{z}_{Bi}}}{1 + \sum_{j \in E} e^{\boldsymbol{\beta}_{Bj} \cdot \mathbf{z}_{Bj}}} \tag{4.5}$$

for any $i \in E = \mathsf{X} \setminus \{B\}$, where time dependence has been suppressed for brevity. The probability of moving from a spike regime to the base regime is described by

$$p_{iB} = \left(1 + e^{-\boldsymbol{\beta}_{iB} \cdot \mathbf{z}_{iB}}\right)^{-1} \tag{4.6}$$

for any $i \in \{S, D\}$. The parameter vector $\boldsymbol{\beta}_{ij}$ and the vector $\mathbf{z}_{ij}$ of exogenous observations have the same length for all $(i, j) \in \mathsf{X}^2$.

## 4.7 Exogenous Variables

The discussion has to this end primarily been about how to improve the fit of the price dynamics to match observed data. The timing of spikes, which is impeded under the assumption of constant transition probabilities, has for this reason not yet been touched upon. The transition mechanism should reflect the fact that prices do follow seasonal trends and that some seasonal components have already been classified. A first step would be to apply a sinusoid that governs transitions. Seasonal cycles may, however, contain cyclical components that are subject to sudden phase shifts, e.g. climate oscillation. Besides, there are well-known indicators with natural connections to spot prices that are worth considering in lieu of simple sinusoidal functions or complex Fourier series. Such indicators are easily incorporated into any MRS models with the logistic functions (4.5) and (4.6). It is therefore worth to examine what exogenous variables influence the soaring prices that are observed in the electricity markets.

The terminology used for such indicators is in this case ambivalent, as they could as well be interpreted as endogenous. Supply and demand in traditional econometric models are obviously directly related to prices, and vice versa, hence from that perspective it would be correct to refer to these terms as endogenous. Our aim, however, is not to model interrelationships, but rather the effect on model calibration when all explanatory variables are fixed.

### 4.7.1 Literature Review

Inhomogeneous transitions in the Markovian framework have a wide range of applications. The very tempting idea to include time-varying transitions in MRS models was realized by Goldfeld and Quandt (1973) and followed up by Diebold et al. (1994) and Filardo (1994) in the case of business cycle durations. Different methods for estimating parameters in this framework were covered by Diebold et al. (1994), Filardo (1994), and Filardo and Gordon (1998). In contrast to the two former studies, which used logistic functions, Filardo and Gordon (1998) let time-varying transitions be modelled using a probit model and presented estimation schemes for MCMC methods. Another example of how the model class can be applied is illustrated by Martinez Peria (2002), who modelled speculative attacks on currencies in the European Monetary System. Yet another example is futures prices of the West Texas Intermediate crude oil which were examined by Chang (2012) using an EGARCH model with a similar regime-switching mechanism.

Electricity prices are arguably more tied to seasonal cycles than businesses, but time-varying transitions in MRS models for such prices still have a modest presence in the literature. Apart from a few exceptions, most models are limited to constant transition probabilities. It is, however, well-known that information about e.g. load or reserve margin helps to forecast spikes.

In a novel approach Davison et al. (2002) developed a hybrid two-state regime-switching model in order to model prices in the American Pennsylvania-New Jersey-Maryland (PJM) market using the supply-demand ratio. The model was later tweaked in a follow-up paper by Anderson and Davison (2008) to model power plant failure.

The study by Fabra and Toro (2005) is the first attempt in the electricity markets that I am aware of that lets transitions of MRS models depend on exogenous variables — in this case demand inter alia — which were governed by logistic functions when studying the Spanish electricity market. A similar technique was used by Mount et al. (2006) and proved to be successful in the sense that load and reserve margin contributed to better spike estimation.

Becker et al. (2007) applied Markov-switching models with time-varying transitions to Australian electricity prices in Queensland. The authors showed that identification of spikes was improved by letting demand patterns and meteorological variables drive transitions between two states. Moreover, they also found that transitions to spike states could be explained by first and foremost load, whereas weather patterns were less significant. Incidentally, they let spikes belong to a beta distribution to better capture extreme price movements.

Following up on those results, Huisman (2008) examined how temperature could be used to detect spikes and showed that spikes in the Dutch APX market were more likely to occur with extreme temperatures by studying average prices during peak hours. In a slightly different approach Rambharat et al. (2005) showed that the price dynamics of the wholesale market in Pennsylvania could be improved by adding temperature effects to a TAR model. Although temperature seems to be a natural factor that is embedded in spot prices, it is not straightforward to incorporate it into an inhomogeneous model. Temperature is somewhat easier to handle in the case of Huisman (2008), since the APX market operates in a relatively small and homogeneous area. For larger areas with greater variety of local temperatures it is more troublesome to extract temperature data into a meaningful yet easy-to-handle proxy. Other variables such as precipitation might as well be more useful to consider depending on what type of climate and power production is typical for a certain area.

Kanamura and Ōhashi (2008) took a different route when they modelled prices in terms of demand and let transition probabilities be functions of current demand and supply as well as trends in demand. By studying daily prices in PJM they found evidence of transition probabilities having a strong connection to seasonality.

With Markov-switching regressions including fundamentals such as demand and excess generation capacity, Karakatsani and Bunn (2008) reached the conclusion that their models were able to predict intraday and day-ahead prices in the UK.

Yet another different approach was taken by Cartea et al. (2009), who showed that estimation benefits from forward-looking information. Rather than working in the Markovian framework, Cartea et al. (2009) approached their modelling of electricity prices in England and Wales with Lévy processes utilizing both mean-reversion and jumps. The underlying price process was in their case driven by two regimes — mimicking one stable and one unstable regime — that were governed by capacity constraints.

A quasi-inhomogeneous model was proposed by Janczura and Weron (2010) and tested with daily base load spot prices for EEX, PJM and New England Power Pool (NEP), where the authors extended the standard time-invariant transitions by estimating periodic transition probabilities for each season of the year. Smoothness was then obtained by

applying a Gaussian kernel to the probabilities. This approach obviously lacks precision on shorter time scales other than on a seasonal basis and does not consider year-on-year variations since it is based upon in-sample data.

Cruz et al. (2011) showed that wind power generation affects Spanish electricity spot prices dramatically due to its ability to penetrate the market as well as its special market position implied by regulations.

In a survey by Løland et al. (2012) the authors found no improvements when including consumption prognoses for forecasting area-specific transmission congestion in Norway. They also considered water reservoir levels, temperature and wind speed but did not implement them in their models, instead they held on to other explanatory variables such as total flow in and out of a specific area.

Christensen et al. (2012) found that abnormal loads influenced both the probability and severity of spikes in four Australian wholesale markets, whereas temperature extremes were found to only impact the occurrence rate of spikes. Similarly, in the two-state SMS model introduced by Eichler and Türk (2013), the authors improved their fit to the same Australian markets with standardized loads.

Zachmann (2013) applied prices of coal, natural gas and $CO_2$-emission allowances in a similar fashion to Karakatsani and Bunn (2008) in order to model prices of the British and German electricity markets. The author found that the model fitted prices in the UK better than Germany, but concluded that demand could improve forecasts and that an extension of the regression to include mean-reversion would be preferable.

Even though two states are not considered to be enough in order to encompass the majority of spikes, most time-varying inhomogeneous models hold on to two states nevertheless and fail to implement higher state order.

### 4.7.2 Variable Selection

A palpable limitation of implementing inhomogeneous models is confined access to real data. Electricity markets have a short history compared with securities and commodities and price quotes do not date back more than roughly ten to twenty years. Moreover, power generation has been reshaped over this time period with the introduction of new sustainable resources. Admittedly, it would be optimal to have more data; if not a decade then at least a handful of years, so that annual meteorological phenomena and extreme outliers would be neutralized. The hourly time scale compensates partly for sampling bias, despite that the study is based upon relatively few years.

From the literature it is clear that demand, generation and reserve margin are inevitable when considering which variables to choose. For the same reasons it is highly plausible to consider forecasts thereof. For other variables to be considered, they have to exhibit some sort of relationship to the market. By looking at the allocation in Figure 1.2 it is tempting to examine relationships between the domestic types of resources and the domestic spot prices. For EEX this means that uncertain production generated by wind turbines and photovoltaic cells are interesting variables. The aggregate data for the Nordic area do not reveal that in particular Denmark has a large portion of electricity

coming from offshore wind parks. Could Danish wind power generation affect the whole Nordic market? More interestingly, could it lead to abrupt price spikes?

A quick overview of the Nordic power generation shows that hydropower is predominant in some countries. Consequently, hydro reservoir levels are arguably another factor that has a close relation to electricity spot prices. In contrast to ephemeral generation such as wind or solar power, which are restricted to current weather conditions, hydro reservoir levels change slowly and function as a long-term component.

Botterud et al. (2010) observed that the difference between spot and futures prices in Nordic wholesale market is linked to some of the above-mentioned variables, such as consumption and reservoir levels.

The following list summarizes some interesting variables that are worth considering when modelling electricity prices in general:

- Power consumption and production are directly related to electricity pricing, cf. Figure 1.1.
- The reserve margin measures the ability of a system to supply peak load. It is defined as the amount of total generation capacity that exceeds annual peak load.
- Wind and solar power generation are known for their dependence on propitious conditions and significant changes in such generation can lead to price spikes.
- The aggregate hydro reservoir level functions as a long-term component which stabilizes prices.

Intuitively, an increase in consumption pushes prices up, whereas the opposite effect is evident for increased production as well as reserve margin, ceteris paribus. Forecasts of as well as relative changes in these variables are also of great interest. In addition to these indicators, a few noteworthy examples to consider are temperature, insolation, transmission capacities, installed capacities, trading volumes, continuous price quotes as well as differences between continuous and day-ahead prices. Intermittent price deviations between areas in the Nordic market occur for instance due to physical transmission constraints.

There are also products outside the electricity market that are affecting power generation. Examples of such products are coal and gas in some countries. Generation that constitute a significant portion of emission allowances is also of interest, as are emission allowances per se, especially EU emission allowances. As mentioned above, temperature and insolation are of course of great interest as well, although rather difficult to implement.

Even though the list of explanatory variables consists of some of the most important indicators, it is far from complete. Two important features of the majority of the above-mentioned indicators are their unambiguity and availability to market participants. In practice, however, the model framework can adopt a wide span of data, but such data may be proprietary and not accessible to everyone, especially market participants.

Chapter 5

# Analysis and Results

## 5.1 Market Data

I have chosen to examine the German and Nordic markets, two markets that have received much attention in the literature and research — that I opt for these two markets is hence not a pure coincidence. First, both markets are among the largest and most liquid markets in the world, which is considered to be of importance, especially when analyzing intraday spikes. Second, the markets constitute a well-balanced mix of energy sources, where the development of sustainable energy has proliferated considering its relatively recent introduction to the markets. On this front, both Denmark and Germany have a large part of their production coming from renewable sources. Wind and solar power are, in particular, two sources that notably stand out since they are heavily reliant on weather conditions. Such uncertainty is prone to cause volatile electricity generation or even abrupt power spikes. This development will probably continue to expand, and with it the risks thereof, thus it is important for any model to capture such behaviour. Finally, despite their geographical proximity, interdependence between EEX and NPS has been shown to be surprisingly low. In a survey covering five substantial markets in Europe, Lindström and Regland (2012) show that EEX experiences spikes at the same time as three of the markets, whereas the Nordic market has a significantly lower interdependence, which in many cases is an order of magnitude. Likewise, an extreme event analysis of both daily and hourly prices will be carried out to investigate this behaviour.

The majority of the data that is used in this thesis has been acquired from European Energy Exchange AG and Nord Pool ASA, whereas forward prices have been obtained from the derivatives markets. Daily as well as hourly day-ahead spot prices are system prices denominated in euro per MWh. The forward price process is taken as the daily settled one-month Phelix Futures on base load for EEX and the daily average of one-month forward prices on base load for NPS, both denominated in euro per MWh. On an hourly scale this means that the forward prices are kept fixed for each day. Since forward prices in the Nordic market have not been denominated in euro before 2006, the study is confined to the beginning of 2006 in order to maintain consistency and avoid modelling currency risk. Missing data are substituted with linearly interpolated values,

while missing boundary values are extrapolated in a similar manner. Splines have also been considered but seem to overfit the data and are therefore discarded. In total only a fraction is missing and the data is therefore deemed very robust.

As illustrated in Figure 5.1, non-stationarity on both markets is heavily reduced by removing the forward-looking information included in the forward prices, though it is still clear that spikes follow seasonal patterns, particularly in the Nordic market, where higher prices are lucidly observed during winters, lower prices are pinpointed to summers and the majority of mean-reverting prices is located in-between. The figure also reveals that the German market is much more volatile than its Nordic counterpart, which on the other hand experiences spikes of less magnitude in addition to frequency, and pronounces more seasonal variations which arguably are related to its resource allocation. After studying the correlation between markets de Jong (2006) asserts that the damping of spikes in NPS can be explained by the relatively large share of hydropower.

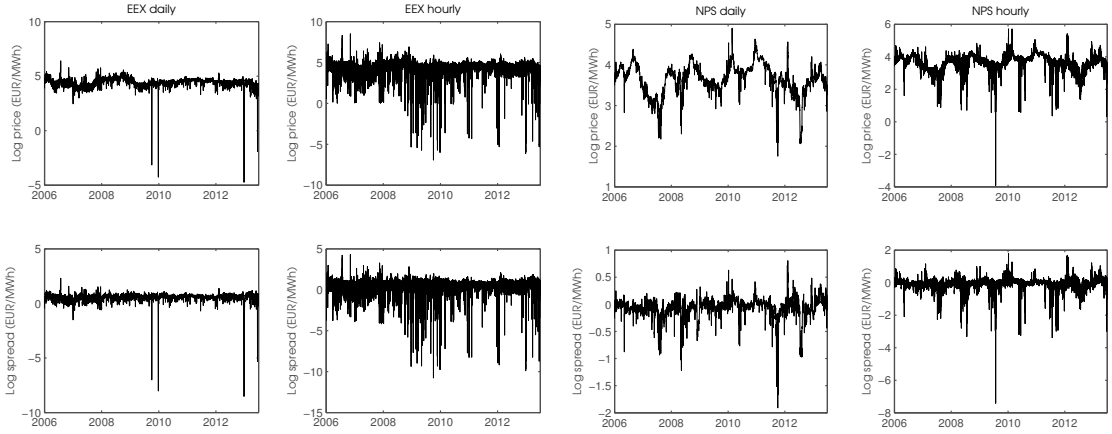



Figure 5.1: The first two columns show transformed daily and hourly day-ahead prices as well as the spreads between transformed prices and transformed one-month futures prices from 1 January 2006 to 30 June 2013 in EEX, where the transformation is taken as the inverse hyperbolic sine. Similarly, prices in NPS are shown in the last two columns with the exception that the spreads are taken as the difference between log prices and one-month log forward prices.

## 5.2 Model Calibration

A word of advice on estimation in MRS models for electricity markets is in place before proceeding with the model calibration. Since an MRS model is a mixture of models with different distributions and underlying parameters, the complete model density will be multimodal and the likelihood function will be affected as a result, hence increasing the difficulty of the optimization problem with respect to the global maximum. Apart from the parameter constraints discussed in Chapter 3, parameter estimation is initialized in

accordance with appropriate initial values with respect to econometrical interpretation as well as earlier research results in order to improve model calibration in the IR setting, since this setting is sensitive to small perturbations of initial parameters values. As an example we have, heuristically, that $\hat{\eta} \approx 1$, which indicates that the forward price process alone contains much of the seasonal trend.

### 5.2.1 Two-State Models

The estimation is initialized with two-state models, although heuristically we believe that three states will be required for improving the goodness of fit, as shown in earlier studies. The introduction of two states is first and foremost to act as a benchmark of further analyses. Like Figure 5.1 suggests, drops seem to be more frequent than up-spikes, but both positive and negative spikes have to be detected nonetheless, in order to obtain accurate forecasts. Since both log-normal and gamma spikes are constrained to one direction only, they will not be able to cover extreme cases in every direction and Gaussian distributions can probably do better in the case of only two states.

Figure 5.2 illustrates identifications of spikes in the German and Nordic markets, two markets that contrast each other. The former exhibits spikes of greater magnitude and is much more volatile, which results in more frequent spikes. Two-state models with drops have better overall fit to real data than up-spikes due to more extreme downward than upward price movements.

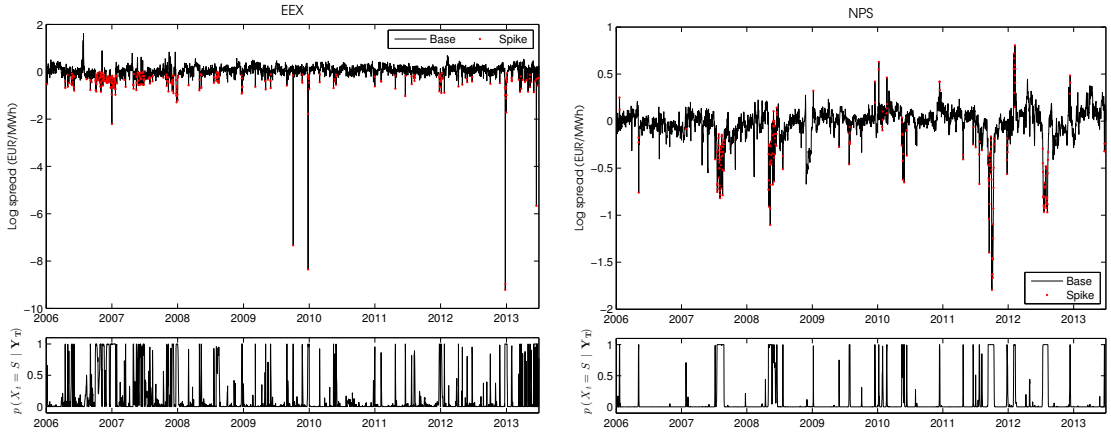



Figure 5.2: Identification of spikes in two-state MRS models for daily system prices in the German and Nordic wholesale markets. Brennan-Schwartz dynamics and log-normal drops fit the German market best, while inverse Brennan-Schwartz dynamics and Gaussian drops best describe the Nordic market. Lower plots illustrate smoothing probabilities with clear identification of volatility clusters.

While calibrations clearly show that EEX exhibits Brennan-Schwartz dynamics and log-normal drops, the log spread in NPS is better described by inverse Brennan-Schwartz dynamics, i.e. $\gamma = -1$, and Gaussian drops. In comparison with NPS, the fit on EEX

differs considerably in the choice of distribution. The reason behind this is the higher variance of the EEX spread, which has not been fully captured by Gaussian spikes. Note also that the two-state model does not detect any up-spikes in the German market due to the capped support bounded by the forward price process, cf. the lower equation of (4.4). There are sufficiently many extreme price quotes that make log-normal drops yield better results. Such up-spikes would, however, be identified by a three-state model, which is a reason for including one more state.

The Nordic market has, by contrast, not experienced the same number of extreme prices and thus a Gaussian mixture performs better in that case. Keep in mind that Gaussian spikes are not restricted to one direction, hence making forecasts suffer, since an up-spike regime can produce extreme positive prices just as well as negative ones. An additional state will have to be introduced in order to cope with this issue.

### 5.2.2 Three-State Models

One distinction between the German and Nordic market lies in the magnitude of the spikes. EEX distinguishes itself from NPS with a wider price range that is better captured by log-normal drops. As shown in the previous section, two states are therefore not enough to capture the dominating frequencies.

Table 5.1: Parameter estimates and log-likelihoods of various three-state MRS models for daily and hourly day-ahead prices in EEX and NPS.

| | Transition probabilities | | | | State parameters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model* | $p_{BS}$ | $p_{BD}$ | $p_{SB}$ | $p_{DB}$ | $\kappa$ | $\eta$ | $\sigma$ | $\gamma$ | $\mu_S$ | $\sigma_S$ | $\mu_D$ | $\sigma_D$ | $\ell$ |
| EEX daily | | | | | | | | | | | | | |
| G-VAS | 0.00 | 0.05 | 0.29 | 0.24 | 0.41 | 1.01 | 0.14 | 0.00 | −1.97 | 3.23 | 0.31 | 0.32 | 737.9 |
| G-CKLS | 0.05 | 0.00 | 0.24 | 0.29 | 0.41 | 1.01 | 0.22 | −0.32 | −0.31 | 0.32 | 1.97 | 3.24 | 737.6 |
| LN-VAS | 0.03 | 0.07 | 0.20 | 0.29 | 0.61 | 1.01 | 0.13 | 0.00 | −1.42 | 0.52 | −1.10 | 0.74 | 744.0 |
| LN-CKLS | 0.02 | 0.06 | 0.19 | 0.30 | 0.60 | 1.01 | 0.40 | −0.75 | −1.41 | 0.51 | -1.08 | 0.74 | 746.5 |
| Γ-VAS | 0.00 | 0.04 | 0.50 | 0.32 | 0.43 | 1.01 | 0.14 | 0.00 | 4.19 | 6.91 | 1.15 | 2.05 | 698.0 |
| Γ-CKLS | 0.02 | 0.05 | 0.15 | 0.28 | 0.57 | 1.01 | 0.61 | −0.99 | 3.23 | 11.79 | 1.19 | 2.29 | 689.5 |
| NPS daily | | | | | | | | | | | | | |
| G-CIR | 0.00 | 0.02 | 0.04 | 0.22 | 0.11 | 1.00 | 0.03 | 0.50 | −0.59 | 0.35 | 0.02 | 0.24 | 3207 |
| G-CKLS | 0.00 | 0.02 | 0.04 | 0.22 | 0.11 | 1.00 | 0.16 | −0.80 | −0.60 | 0.35 | 0.01 | 0.25 | 3230 |
| LN-CIR | 0.01 | 0.01 | 0.19 | 0.14 | 0.13 | 1.00 | 0.03 | 0.50 | −2.05 | 0.81 | −0.97 | 0.69 | 3254 |
| LN-CKLS | 0.01 | 0.01 | 0.13 | 0.12 | 0.16 | 1.00 | 0.37 | −1.46 | −1.77 | 0.67 | −1.09 | 0.72 | 3227 |
| Γ-VAS | 0.01 | 0.01 | 0.19 | 0.14 | 0.13 | 1.00 | 0.06 | 0.00 | 1.58 | 8.16 | 2.30 | 4.75 | 3267 |
| Γ-CIR | 0.01 | 0.01 | 0.20 | 0.15 | 0.13 | 1.00 | 0.03 | 0.50 | 1.54 | 8.25 | 2.28 | 4.77 | 3250 |
| Γ-CKLS | 0.01 | 0.01 | 0.13 | 0.12 | 0.16 | 1.00 | 0.38 | −1.48 | 2.46 | 11.47 | 2.22 | 5.02 | 3234 |
| EEX hourly | | | | | | | | | | | | | |
| G-CKLS | 0.00 | 0.03 | 0.09 | 0.13 | 0.07 | 1.02 | 0.05 | 0.48 | −1.55 | 1.80 | 0.08 | 0.47 | 23564 |
| LN-CKLS | 0.02 | 0.02 | 0.18 | 0.12 | 0.10 | 1.01 | 0.11 | 0.01 | −1.00 | 0.57 | −0.46 | 0.83 | 29206 |
| Γ-CKLS | 0.02 | 0.02 | 0.18 | 0.13 | 0.09 | 1.01 | 0.11 | −0.01 | 2.63 | 6.46 | 1.22 | 1.30 | 28205 |
| NPS hourly | | | | | | | | | | | | | |
| G-CKLS | 0.03 | 0.01 | 0.09 | 0.05 | 0.02 | 1.01 | 0.02 | 0.43 | 0.02 | 0.13 | 0.46 | 0.57 | 88980 |
| LN-CKLS | 0.02 | 0.01 | 0.11 | 0.06 | 0.04 | 1.01 | 0.03 | −0.01 | −1.84 | 0.78 | −1.17 | 0.95 | 97267 |
| Γ-CKLS | 0.02 | 0.01 | 0.11 | 0.07 | 0.04 | 1.01 | 0.03 | −0.01 | 1.84 | 9.07 | 1.27 | 2.71 | 97576 |

*Models are denoted by spike distribution followed by base regime dynamics, where G, LN and Γ denote the Gaussian, log-normal and gamma distributions, respectively.

Calibration of parameters of three-state models for daily and hourly day-ahead prices in EEX and NPS is presented in Table 5.1. The results are in parity with earlier studies, see e.g. Lindström and Regland (2012), who also found similar parameter values such

as for instance $\hat{\eta} \approx 1$. In the cases of daily prices the CKLS dynamics try to capture the leverage effect instead of its inverse counterpart. Other dynamics, such as Vašíček, CIR and Brennan-Schwartz, have been tested for this reason, though all combinations are not presented due to not improving the overall fit. Notwithstanding that none of these dynamics should be able to achieve a higher likelihood than CKLS dynamics, some calibrations attain slightly better likelihoods nevertheless. This indicates that the parameter that invokes heteroskedasticity increases the computational complexity more than necessary, thus impeding parsimony. A negative $\gamma$ might also indicate a preference for Vašíček dynamics, since closed-form solutions are not available in such cases and parsimony is preferable.

The inverse leverage effect is evident when Gaussian spikes drive hourly prices. On the other hand, calibration of hourly prices with log-normal and gamma spikes suggests that Vašíček dynamics are enough. This does not necessarily mean that the models are unable to encompass the inverse leverage effect for log-normal and gamma spikes, but it rather indicates that the effect is already inherent in the specific spike regimes. The results also evince that some cases of three states with Gaussian spikes are not needed, since transitions from the base regime to one of the spike regimes rarely occur.

Not only are the German and Nordic markets characterized by distinct spikes, but the length of volatility clusters is also remarkably different. From the calibration it is clear that the probability of remaining in the drop regime, cf. Table 5.1, is significantly smaller for EEX, irrespective of the time scale. For daily non-Gaussian spikes in NPS the probability of leaving the drop state is more than twofold. Daily mean-reverting prices and hourly spikes and drops in the Nordic market are also more enduring, while periods of daily up-spikes and hourly mean-reverting prices are of approximately the same length in both markets.

The introduction of an additional state improves the log-likelihood of all models, cf. Table 5.2. LR tests and BIC-values, which address the problem of overfitting data by using too many parameters, clearly show that the additional spike state contributes significantly to the improvement of the overall model. Those tests, however, do not consider individual regimes, hence K-S tests are carried out for each spike state. The results show that daily system prices in such states in NPS are captured by the model parameters. The goodness of fit is not as good for the daily EEX spread mostly due to the difficulty in fitting extreme downward movements. For this reason it would be preferable to examine different families of distributions in order to find an optimal mix of spike distributions that could yield a higher goodness of fit.

Hourly price models do not perform as well under the K-S test as the corresponding ones for daily prices. This is probably due to the enlarged sample which implies that the supremum distance must be substantially smaller in order to not reject the null hypothesis, which states that the model-implied distribution is the empirical one. As proposed by Janczura and Weron (2010), the Cramér-von Mises test, which is based upon the $L^2$ norm, could probably be a better test statistic as it considers the entire cdf instead of the supremum distance. Another reason might be that the deseasonalization has not fully removed the intraday trend such that the identification of spikes is less

Table 5.2: Log-likelihoods, BIC-values and $p$-values of LR and Kolmogorov-Smirnov tests for various three-state MRS models for daily day-ahead prices in EEX and NPS. The LR tests are carried out for analogous two- and three-state models. Corresponding values for two-state models with drops are given in parentheses. Bold font emphasizes $p$-values of LR tests less than 0.05 and $p$-values of K-S tests greater than 0.05.

| | | | | Kolmogorov-Smirnov | |
|---|---|---|---|---|---|
| Model* | $\ell$ | BIC | LR | Spike | Drop |
| EEX daily | | | | | |
| G-VAS | 737.9 (577.1) | −1389 (−1091) | **0.0000** | **0.1235** | 0.0000 |
| G-CKLS | 737.6 (590.4) | −1380 (−1117) | **0.0000** | 0.0000 | **0.1210** |
| LN-VAS | 744.0 (665.9) | −1401 (−1269) | **0.0000** | **0.1313** | 0.0000 |
| LN-CKLS | 746.5 (679.5) | −1398 (−1296) | **0.0000** | **0.1343** | 0.0000 |
| Γ-VAS | 698.0 (623.9) | −1309 (−1185) | **0.0000** | **0.5042** | 0.0000 |
| Γ-CKLS | 689.5 (633.0) | −1284 (−1203) | **0.0000** | **0.4103** | 0.0000 |
| NPS daily | | | | | |
| G-CIR | 3207 (3145) | −6327 (−6234) | **0.0000** | 0.0432 | **0.5021** |
| G-CKLS | 3230 (3148) | −6364 (−6232) | **0.0000** | **0.0782** | **0.5990** |
| LN-CIR | 3254 (3067) | −6421 (−6234) | **0.0000** | **0.6192** | **0.1327** |
| LN-CKLS | 3227 (3066) | −6358 (−6069) | **0.0000** | **0.2150** | **0.5080** |
| Γ-VAS | 3267 (3063) | −6440 (−6064) | **0.0000** | **0.8366** | **0.3250** |
| Γ-CIR | 3250 (3071) | −6413 (−6079) | **0.0000** | **0.7433** | **0.3960** |
| Γ-CKLS | 3234 (3071) | −6372 (−6078) | **0.0000** | **0.4338** | **0.4209** |

*Models are denoted by spike distribution followed by base regime dynamics, where G, LN and Γ denote the Gaussian, log-normal and gamma distributions, respectively.

efficient. With this in mind, a graphical comparison of the empirical and model-implied distributions highlighting the overall goodness of fit is in order. Probability distributions of all regimes are shown in Figures 5.3 and 5.4, which illustrate how the shape, especially the skewness and kurtosis, of the empirical data is captured by the models. Above all, the theoretical distributions fit well for all empirical distributions, regardless of market and time scale. Probability distributions of the complete models are illustrated in Figure 5.5 and perspicuously distinct extreme prices from the mean-reversion level. The tails of negative prices compared with positive ones are fatter as expected due to the greater magnitude of negative prices, particularly in the German market.

The calibration of three-state models can finally be crystalized into optimal base dynamics and spike distributions for every market and time scale. The choice of spike distribution can in most cases be determined by comparing likelihoods of different models. Additional adequate methods for model selection strengthen the choice when a mere comparison of the likelihoods is indefinite, which is the case of daily prices in the Nordic market. Like Lindström and Regland (2012) conclude for daily prices in NPS, the results interestingly show that Vašíček dynamics and gamma spikes also perform best for hourly prices in NPS. Although other dynamics, such as CIR, also yield good results, Vašíček dynamics seem to do slightly better when comparing distribution plots. As a result we conclude that Vašíček dynamics with gamma spikes are preferable for both daily and

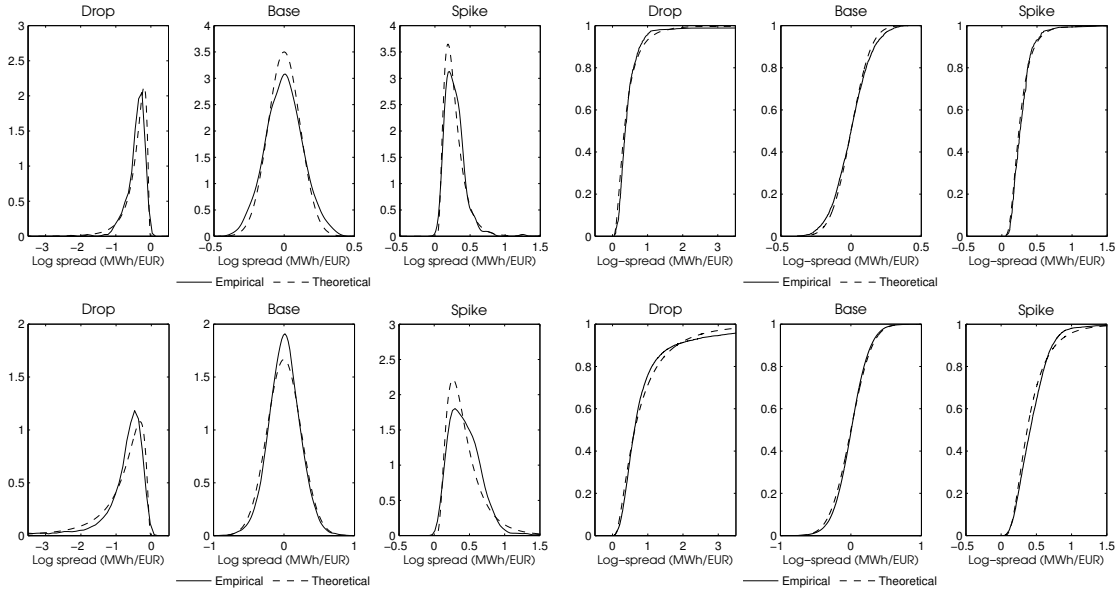



Figure 5.3: Empirical and theoretical pdfs and cdfs of each regime of three-state MRS models with Vašíček dynamics and log-normal spikes for daily (upper row) and hourly (lower row) log spreads in EEX.

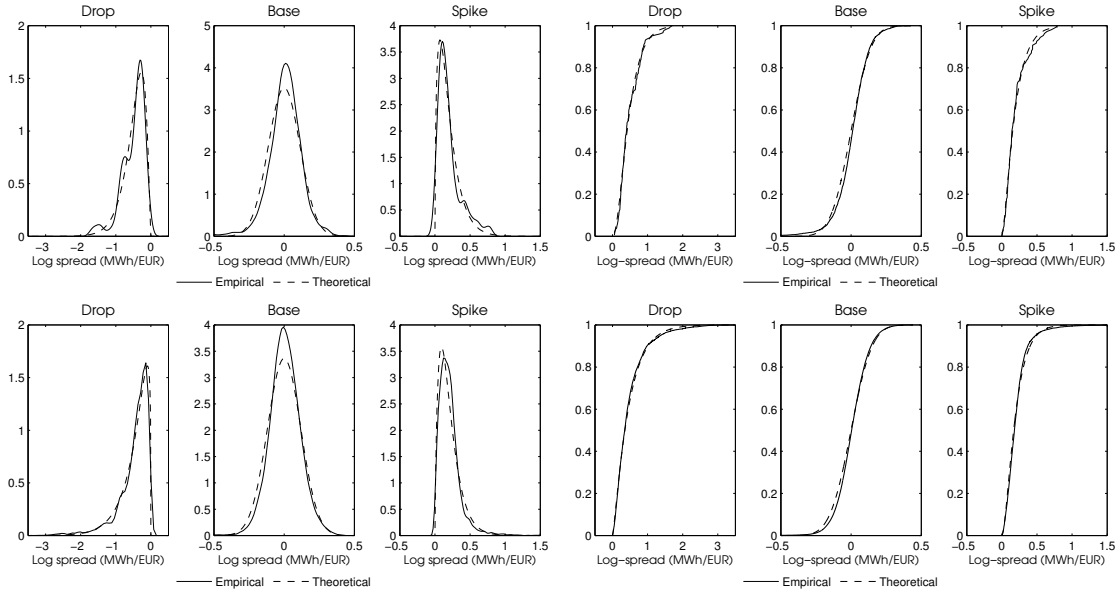



Figure 5.4: Empirical and theoretical pdfs and cdfs of each regime of three-state MRS models with Vašíček dynamics and gamma spikes for daily (upper row) and hourly (lower row) log spreads in NPS.

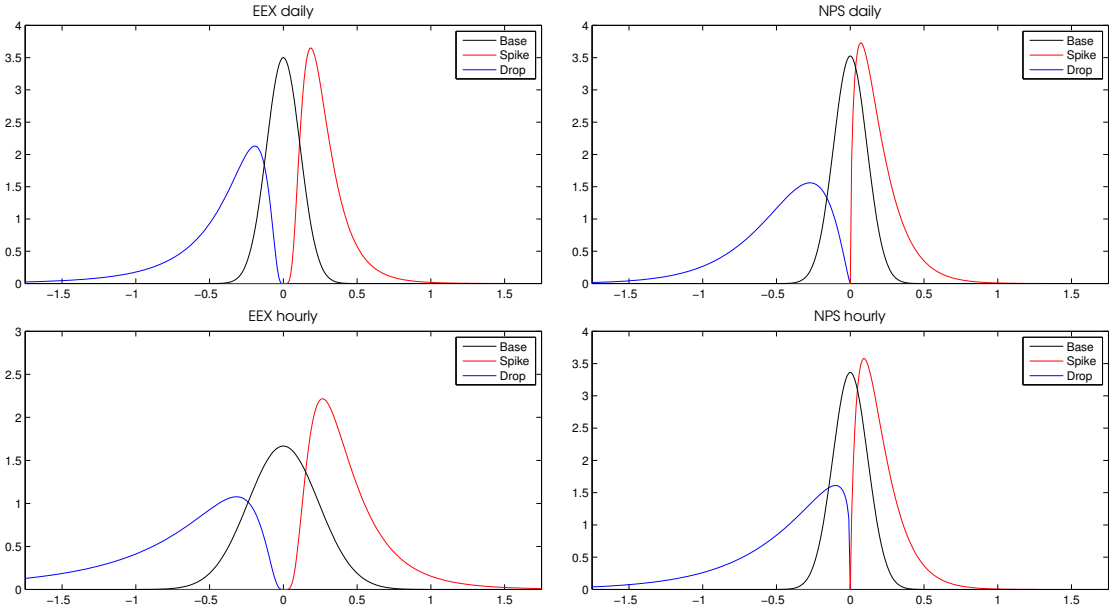



Figure 5.5: Probability distributions of all regimes in the German and Nordic markets, where Vašíček dynamics together with log-normally and gamma distributed spikes, respectively, best describe price movements.

hourly prices in NPS, whereas Vašíček dynamics with log-normal spikes are the optimal choice for EEX, irrespective of the time scale.

In contrast to the imperfect identification of spikes in two-state models, cf. Figure 5.2, the introduction of an additional regime results in three-state models outperforming the bistate models. The calibration of three-state models, as illustrated in Figure 5.6, clearly shows that up-spikes are captured by the models. The difference is substantial, especially for daily prices in EEX where spikes are more frequent and where a Gaussian mixture with two states is not the optimal choice. The characteristic state transitions in both markets are also highlighted in Figure 5.6; drops in the Nordic market are in particular longer-lasting.

## 5.3 Inhomogeneous Models

Most of the exogenous variables presented in the previous chapter are considered in this thesis. Consumption (C), consumption prognosis (CP), production (P), production prognosis (PP), reserve margin reserve (RM), reserve margin prognosis (RMP), Danish wind power production (W) and Danish wind power production prognosis (WP) will be used to study NPS. Power consumption and generation data for the Nordic countries are available from the beginning of 2000, except for Denmark where measurements are available from 1 October 2000, whereas forecasts of both consumption and production are available from 3 March 2008 and 1 July 2010, respectively. Settled and forecasted

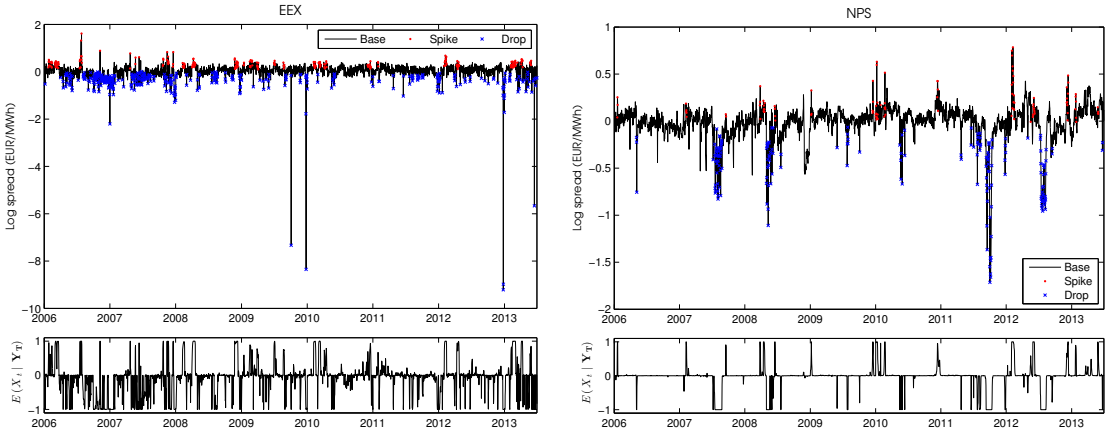



Figure 5.6: Identification of spikes in three-state MRS models for daily day-ahead system prices in EEX (left panel) and NPS (right panel). Vašíček dynamics together with log-normal and gamma spikes fit the German and Nordic market best. Lower plots illustrate the conditional expectation of the state process, where 1 corresponds to spike, 0 to base and $-1$ to drop. Note the clear identification of volatility clusters.

wind power production are available from 14 September 2009. For this reason several estimations will be carried out in order to compare all variables with each other for identical time periods. This is also done since some exogenous data only cover a few years and the sample sizes are quite small, particularly for daily prices.

All determinants, except wind power production, are aggregate data for Denmark, Finland, Norway and Sweden. The Baltic states are left out since they are relatively small and were just recently introduced on NPS. Wind power data for more areas would be of interest, but is for this reason excluded. It is worth stressing that future research including those as well as other areas would be covetable in order to better understand the similarities and differences between areas. Reserve margin data have not been available; instead a slightly modified definition of the reserve margin will be used. Rather than the installed capacity and annual peak load, we use the instantaneous consumption and production as proxies, so that we still measure the excess generation capacity of the system. Similarly, reserve margin forecasts are based upon predicted supply and demand. Water reservoir levels are discarded since such figures are quoted only once a week and too much information would probably be lost in order to consider it as an indicator for daily or even hourly prices. In comparison with the above-mentioned indicators, water reservoir levels are the least volatile indicator and unlikely to capture spikes as good as the other determinants. Not to mention, water reservoir levels do incorporate long-term seasonal patterns and for models including a relatively large number of indicators it would presumably be a good idea to also include water reservoir levels in order to capture such low frequencies.

Similar indicators have been considered for EEX, but data for those as well as solar

power have unfortunately not been available. More interestingly, the installed capacity in Germany allows for other indicators such as gas prices and EU carbon emission allowances (EUAs); both are taken as daily settlement prices denominated in euro per MWh and EUA, respectively. The natural gas spot market in Germany is divided into a couple of market areas, each with its own settlement price, thus the gas price is taken as the mean of all market areas. Since both gas prices and EUAs are settled once per day, the settlement prices are fixed for all hours of each day when modelling hourly prices. Finally, all exogenous inputs, which are henceforth denoted by $\tilde{Z}$, are normalized with respect to their maxima in order to obtain comparable and consolidated results.

### 5.3.1 Model Calibration and Its Implications

The inhomogeneous models presented henceforth have all been subject to the same tests as reported in Table 5.2, i.e. the LR and K-S tests as well as the Bayesian information criterion. Although not reported herein, the tests could not reject any model at a 95% confidence level. Logistic coefficients of three-state models for daily and hourly prices are listed in Tables 5.3 and 5.4. Subscripts of the beta coefficients indicate which transitions and variables they represent. The coefficient $\beta_{BS,0}$ represents for example the weight of the intercept of the multinomial logistic function for transitions from the base to the spike regime, whereas $\beta_{BS,1}$ is the weight of the exogenous variable of the same logistic function. The quadratic term, $\tilde{z}^2$, has also been added as a third parameter to investigate its implication on transitions. However, the experience is that the variance escalates most of the times as the number of parameters increases, thus yielding a small number of significant coefficients. The same occurs when simultaneously combining a couple of exogenous variables, such as e.g. demand and generation. Quadratic and cross terms are, as a consequence, omitted from the estimations and only one exogenous variable is considered for each calibration.

Interestingly, a large number of the coefficients, in particular those affecting exogenous variables, are statistically significant, hence confirming our a priori believes that especially consumption and production ought to have a close relationship to spot prices. All significant coefficients for daily prices in the Nordic market are positive for transitions to up-spikes and negative for transitions to down-spikes. Loosely speaking, this means that an increase (decrease) in the exogenous variable consequently increases (decreases) the probability of spikes (drops).

As the results clearly show, the number of significant coefficients varies with the time period, though the results are consistent overall. While the calibration results only yield significant exogenous coefficients when including EUAs for the spot price in EEX, the Nordic market has a closer relationship to the tested data. The results show that consumption, production and forecasts thereof are significant for the transitions form the base regime. While production data from 2006 and 2008 are able to capture the transitions from up-spikes to mean-reverting prices, consumption data from 2006 results in all coefficients being significant. The reserve margin is significant for transitions from both up-spikes and down-spikes, but not for transition from the base state. The reserve margin prognosis does not yield any significant coefficients of interest. The coefficients

that are representing wind power and forecasts thereof are statistically significant for transitions to up-spikes. While their signs are positive for daily prices, i.e. an increase in wind power production increases the probability of up-spikes, the signs are negative for hourly prices.

Table 5.3: Exogenous data coefficients of three-state MRS models using Vašíček dynamics together with log-normal and gamma spikes for daily day-ahead prices in EEX and NPS, respectively. Significant coefficients are emphasized in bold font with corresponding $p$-values printed beneath in parentheses. All time series are evaluated from the starting date specified below each market until 30 June 2013.*

| Data | $\beta_{BS,0}$ | $\beta_{BS,1}$ | $\beta_{BD,0}$ | $\beta_{BD,1}$ | $\beta_{SB,0}$ | $\beta_{SB,1}$ | $\beta_{DB,0}$ | $\beta_{DB,1}$ |
|---|---|---|---|---|---|---|---|---|
| EEX-Gas | **−2.94** | −1.15 | **−3.40** | 1.20 | −1.10 | 2.96 | −0.13 | 0.51 |
| 2008-Q1 | (0.006) | (0.587) | (0.000) | (0.163) | (0.519) | (0.434) | (0.855) | (0.696) |
| EEX-Gas | **−3.13** | −0.29 | **−3.27** | 1.11 | −3.46 | 10.59 | 0.07 | 0.19 |
| 2009-Q1 | (0.002) | (0.885) | (0.000) | (0.208) | (0.306) | (0.343) | (0.926) | (0.885) |
| EEX-EUA | **−1.53** | **−2.70** | **−1.09** | **−2.69** | 4.22 | −5.55 | −0.14 | 0.67 |
| 2009-Q1 | (0.035) | (0.043) | (0.003) | (0.000) | (0.252) | (0.227) | (0.804) | (0.496) |
| NPS-C | **−27.84** | **28.14** | **6.53** | **−17.54** | **13.27** | **−16.34** | **−6.02** | **7.71** |
| 2006-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.020) | (0.013) | (0.002) | (0.013) |
| NPS-P | **−30.96** | **31.23** | **5.87** | **−16.75** | **29.80** | **−33.61** | **−4.31** | 4.42 |
| 2006-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.019) | (0.018) | (0.036) | (0.177) |
| NPS-RM | **−4.45** | 1.11 | **−4.30** | 0.08 | **−1.79** | 0.45 | **−1.63** | −1.48 |
| 2006-Q1 | (0.000) | (0.059) | (0.000) | (0.853) | (0.000) | (0.474) | (0.000) | (0.089) |
| NPS-C | **−31.02** | **31.84** | **7.32** | **−19.04** | 10.95 | −13.83 | **−4.96** | 6.61 |
| 2008-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.110) | (0.076) | (0.018) | (0.057) |
| NPS-CP | **−31.61** | **32.70** | **7.43** | **−19.43** | 14.07 | −17.34 | **−4.79** | 6.36 |
| 2008-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.128) | (0.100) | (0.014) | (0.050) |
| NPS-P | **−39.98** | **41.52** | **5.49** | **−15.64** | **29.07** | **−32.42** | −3.58 | 3.92 |
| 2008-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.022) | (0.019) | (0.090) | (0.245) |
| NPS-RM | **−4.20** | 0.68 | **−4.22** | −0.79 | **−1.91** | 0.47 | **−1.41** | −1.04 |
| 2008-Q4 | (0.000) | (0.173) | (0.000) | (0.074) | (0.000) | (0.448) | (0.000) | (0.256) |
| NPS-C | **−25.75** | **25.64** | **11.89** | **−27.26** | 52.51 | −59.55 | **−5.88** | 7.96 |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.078) | (0.076) | (0.034) | (0.083) |
| NPS-CP | **−24.69** | **24.84** | **11.41** | **−26.71** | 44.76 | −51.01 | **−6.26** | 8.65 |
| 2009-Q4 | (0.000) | (0.001) | (0.000) | (0.000) | (0.089) | (0.087) | (0.034) | (0.081) |
| NPS-P | **−44.66** | **46.24** | **8.77** | **−21.54** | 110.46 | −126.18 | −0.22 | −1.58 |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.288) | (0.287) | (0.924) | (0.681) |
| NPS-RM | **−4.52** | 0.54 | **−4.38** | −0.43 | **−1.83** | **−2.67** | −0.42 | **−3.96** |
| 2009-Q4 | (0.000) | (0.442) | (0.000) | (0.450) | (0.000) | (0.001) | (0.232) | (0.005) |
| NPS-W | **−4.58** | 1.66 | **−4.52** | 1.20 | **−3.96** | **5.37** | −0.48 | −4.43 |
| 2009-Q4 | (0.000) | (0.101) | (0.000) | (0.335) | (0.000) | (0.003) | (0.298) | (0.079) |
| NPS-WP | **−4.59** | 1.71 | **−4.58** | 1.41 | **−3.98** | **5.49** | −0.58 | −3.68 |
| 2009-Q4 | (0.000) | (0.079) | (0.000) | (0.234) | (0.000) | (0.002) | (0.206) | (0.074) |
| NPS-C | **−40.08** | **40.17** | **12.77** | **−29.42** | 13.92 | −18.56 | −3.28 | 2.54 |
| 2010-Q4 | (0.010) | (0.014) | (0.001) | (0.000) | (0.261) | (0.196) | (0.341) | (0.661) |
| NPS-CP | **−34.72** | **34.92** | **10.37** | **−25.13** | 25.17 | −28.83 | −3.22 | 2.41 |
| 2010-Q4 | (0.004) | (0.007) | (0.004) | (0.000) | (0.083) | (0.078) | (0.335) | (0.668) |
| NPS-P | **−41.02** | **41.73** | **13.19** | **−29.59** | 135.87 | −154.71 | −0.38 | −2.41 |
| 2010-Q4 | (0.002) | (0.004) | (0.000) | (0.000) | (0.097) | (0.100) | (0.913) | (0.674) |
| NPS-PP | **−46.02** | **47.29** | **12.71** | **−29.33** | 35.09 | −38.46 | −1.22 | −1.01 |
| 2010-Q4 | (0.004) | (0.006) | (0.001) | (0.000) | (0.166) | (0.165) | (0.723) | (0.860) |
| NPS-RM | **−5.45** | 0.81 | **−4.64** | 0.69 | −1.67 | −2.25 | **−1.00** | −3.04 |
| 2010-Q4 | (0.000) | (0.739) | (0.000) | (0.364) | (0.234) | (0.746) | (0.048) | (0.094) |
| NPS-RMP | **−5.32** | 1.34 | **−4.55** | 0.55 | −2.23 | −3.33 | **−1.82** | −2.42 |
| 2010-Q4 | (0.000) | (0.600) | (0.000) | (0.550) | (0.103) | (0.647) | (0.000) | (0.247) |
| NPS-W | **−3.52** | −5.93 | **−4.64** | 0.26 | **−8.43** | **12.48** | −0.85 | −6.95 |
| 2010-Q4 | (0.000) | (0.131) | (0.000) | (0.885) | (0.000) | (0.001) | (0.210) | (0.148) |
| NPS-WP | **−4.80** | **2.57** | **−4.70** | 0.26 | **−7.19** | **11.16** | −0.88 | −6.24 |
| 2010-Q4 | (0.000) | (0.008) | (0.000) | (0.887) | (0.000) | (0.001) | (0.180) | (0.133) |

*Statistically significant parameters at a 95% confidence level using the Wald statistic $W = \hat{\theta} \big/ \sqrt{\mathbb{V}(\hat{\theta})}$.

Calibrations for hourly data yield more significant coefficients en masse compared with daily prices, especially for the German market where almost all transition coefficients are significant, except for most transitions to and from up-spikes. A large number of significant coefficients is also evident for the Nordic market. While shorter time horizons in general yield less significant number of coefficients for daily prices, the same pattern

Table 5.4: Exogenous data coefficients of three-state MRS models using Vašíček dynamics together with log-normal and gamma spikes for hourly day-ahead prices in EEX and NPS, respectively. Significant coefficients are emphasized in bold font with corresponding $p$-values printed beneath in parentheses. All time series are evaluated from the starting date specified below each market until 30 June 2013.*

| Data | $\beta_{BS,0}$ | $\beta_{BS,1}$ | $\beta_{BD,0}$ | $\beta_{BD,1}$ | $\beta_{SB,0}$ | $\beta_{SB,1}$ | $\beta_{DB,0}$ | $\beta_{DB,1}$ |
|---|---|---|---|---|---|---|---|---|
| EEX-Gas | **−6.46** | **1.76** | **−3.42** | **−1.85** | **−1.01** | −0.22 | **−0.92** | **−1.05** |
| 2008-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.545) | (0.000) | (0.000) |
| EEX-Gas | **−5.18** | **−1.60** | **−3.38** | **−2.01** | **−1.06** | −0.23 | **−0.69** | **−1.90** |
| 2009-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.612) | (0.000) | (0.000) |
| EEX-EUA | **−5.86** | −0.03 | **−5.35** | **1.26** | **−0.86** | −0.47 | **−2.50** | **1.37** |
| 2009-Q1 | (0.000) | (0.895) | (0.000) | (0.000) | (0.000) | (0.170) | (0.000) | (0.000) |
| NPS-C | **−32.22** | **34.12** | **2.07** | **−10.81** | **2.00** | **−3.64** | **−3.62** | **2.87** |
| 2006-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-P | **−30.99** | **31.72** | **2.64** | **−12.14** | **4.30** | **−6.35** | **−3.62** | **3.86** |
| 2006-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-RM | **−12.93** | **−4.17** | **−9.42** | **0.81** | **−0.19** | **0.81** | **−1.53** | **−2.40** |
| 2006-Q1 | (0.000) | (0.000) | (0.000) | (0.000) | (0.006) | (0.000) | (0.000) | (0.000) |
| NPS-C | **−32.03** | **33.54** | **6.49** | **−21.07** | **3.12** | **−5.00** | **−1.47** | **−1.59** |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-CP | **−31.10** | **33.02** | **2.04** | **−11.47** | **1.81** | **−3.65** | **−3.03** | **2.18** |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-P | **−34.97** | **36.95** | **4.57** | **−15.88** | **2.63** | **−4.85** | **−4.83** | **5.04** |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-RM | **−12.25** | **−1.60** | **−9.97** | **−0.74** | **−1.56** | **−2.98** | **−1.79** | **−1.15** |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-W | **−11.27** | **−0.95** | **−10.40** | **3.83** | **0.86** | **−9.78** | **−2.10** | **−0.68** |
| 2009-Q4 | (0.000) | (0.009) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.047) |
| NPS-WP | **−11.16** | **−2.25** | **−10.46** | **3.80** | **−0.58** | −0.66 | **−2.15** | −0.48 |
| 2009-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.073) | (0.000) | (0.136) |
| NPS-C | **−29.87** | **30.84** | **3.41** | **−13.24** | **1.66** | **−3.43** | **−4.92** | **5.03** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) |
| NPS-CP | **−31.58** | **33.14** | **3.40** | **−13.57** | **1.69** | **−3.56** | **−3.49** | **2.02** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.006) |
| NPS-P | **−34.05** | **35.95** | **4.85** | **−16.26** | **3.04** | **−5.03** | **−2.97** | 1.13 |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.158) |
| NPS-PP | **−32.39** | **34.15** | **4.97** | **−16.74** | **1.83** | **−3.54** | **−4.46** | **3.92** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) |
| NPS-RM | **−11.09** | **−2.12** | **−10.00** | **−0.63** | **−1.01** | **−3.42** | **−2.08** | **−1.26** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-RMP | **−11.95** | **−6.78** | **−10.65** | **−1.93** | **−0.71** | **0.66** | **−2.37** | **−1.82** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) |
| NPS-W | **−11.65** | **7.90** | **−10.58** | **7.50** | **0.37** | **−5.76** | **−2.94** | **2.03** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NPS-WP | **−10.77** | **5.30** | **−10.53** | **6.15** | **−0.63** | **−0.72** | **−2.25** | **−0.74** |
| 2010-Q4 | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.002) | (0.000) | (0.048) |

*Statistically significant parameters at a 95% confidence level using the Wald statistic $W = \hat{\theta} \big/ \sqrt{\mathbb{V}(\hat{\theta})}$.

is not observable for hourly prices. This is presumably due to fewer transitions that occur during shorter time periods. The coefficients of the transitions from spikes to the base regime for daily prices are in particular affected in this way. A substantially larger number of transitions taking place intraday could explain the varying number of significant coefficients between the two time scales.

Intercept coefficients are not of special interest, but it is rather their proportion to the exogenous coefficients that is of importance. It is, however, hard to tell if spikes are detected by the exogenous data by only considering whether a coefficient is significant or not. Graphical representations of calibration results for a selected set of estimations with large numbers of significant variables are for this reason added in Appendix A in order to enhance the interpretation of the coefficients. The disposition of the plots is as follows: Similar to some of the precedent figures showing identification of spikes, the deseasonalized spread and conditional expectation of the state process are depicted in the two upper

rows. The middle row, which shows the normalized exogenous process, is followed by time-varying transition probabilities, where the penultimate row illustrates transition probabilities from the base regime and the final row displays transition probabilities from the spike regimes. Blue and red solid lines represent transitions to up-spikes and down-spikes in the fourth row and transitions from up-spikes and down-spikes in the fifth row, respectively. The notation $p_{ij}(t)$ in the figures is used as shorthand notation for $p_{ij}(\tilde{\mathbf{z}}_t, \boldsymbol{\beta})$.

The gas and EUA data in Figure A.1 show no typical relation to spikes and the corresponding transition probabilities do not improve the timing of spikes. For this reason we cannot improve the three-state model with only such information. This finding is in line with de Jong and Schneider (2009), who find that Dutch gas and electricity spot prices lack a positive relationship since such a relationship is already captured by the forward market.

Three-state inhomogeneous Markov-switching models, however, prove to be pivotal when modelling daily and hourly power spikes in the Nordic market. The models are able to capture the occurrence of spikes by including normalized consumption and production, see Figure A.2. Both panels depict very typical seasonal cycles, and the exogenous variables are able to time the transitions to spikes from the base regime very clearly. Intuitively, we expect higher prices during the winter season. Not only does the probability of moving to the base regime from up-spikes decrease during such periods, but transitions to the base regime from the drop regime are also at the same time more likely. Since one-step transitions between up-spikes and down-spikes are not allowed, this means that the model compensates the probability of reaching an up-spike by pushing transitions in that direction in the sense that transitions from drops to mean-reverting prices are more frequent than remaining in the drop state. The effect is a consequence of the shape of the exogenous process and the signs of the exogenous coefficients $\beta_{SB,1}$ and $\beta_{DB,1}$ in Table 5.3. Given the consumption profile, a sign of $\beta_{SB,1}$ that is different from $\beta_{DB,1}$ results in such behaviour. The coefficient that is representing transitions from the drop to the base state driven by power production is not significant and variations in the transition probabilities are, as a consequence, not as clear since they lack greater amplitude compared with the corresponding probabilities in the left panel. The above-mentioned effect of pushing transitions in one direction during certain seasons will for this reason not be as prominent. Note also that transition probabilities for up-spikes tend to be higher than for down-spikes. Above all, the great deviations in the transition probabilities result in more frequent spikes during the summer and winter seasons. This is in accordance with our intuition, that is high prices are expected during cold periods.

Similar seasonal cycles are obtained with forward-looking information, see Figure A.3. Note that production forecasts yield higher transition probabilities in-sample with more extreme observations. Like the measured data, the production forecasts are, however, not able to capture the transitions from drops to mean-reverting prices. In fact, the probabilities of moving to the base regime are less accurate, which is probably due to the smaller sample size. Consumption forecasts do, by contrast, capture the seasonal cycles of all transition probabilities. Like measured consumption, they also generate a relatively

small probability of experiencing drops. Even though the results in Table 5.3 do not yield significance of all coefficients for consumption prognosis, the pattern in Figure A.3 is so clear that it is hard to neglect, since the implication of constant transition probabilities would be a static model. Notice also that the Wald test produces a $p$-value of the $\beta_{DB,1}$ coefficient that is a borderline case.

The left panel in Figure A.4 shows the results for the reserve margin, which is a significant indicator for transitions from up-spikes and down-spikes. Spikes and drops are poorly timed by the reserve margin and the effect of pushing up-spikes to down-spikes during some periods is not evident as in the case of consumption, cf. the signs of the coefficients $\beta_{SB,1}$ and $\beta_{DB,1}$ in Table 5.3. The absence of the effect is in part due to $\beta_{SB,1}$ and $\beta_{DB,1}$ are having the same sign. Above all, the seasonal trends are not as typical as for consumption and production.

The results using wind power generation are illustrated in Figure A.5. Contrary to our intuition, an increase in wind power generation forecasts sometimes seems to increase the probability of up-spikes, though ramping up wind power production could lead to higher prices if the generation suddenly diminishes. Transitions that follow actual wind power production carry more noise and do not distinguish states well.

To this end the discussion about timing of spikes has solely considered daily prices in NPS and while the results for hourly prices are similar there remain some distinctions. While a larger amount of the coefficients are significant, the graphical inspection reveals that the calibrations are less accurate, for consumption and production see Figure A.6, and for prognoses thereof see Figure A.7. The effect of pushing transitions in one direction during some seasons is not as clear, since the amplitude of the transition probabilities is not as great and the probabilities do not deviate as much. Like for the daily data, transition probabilities to up-spikes are in general higher than to down-spikes.

The hourly reserve margin, see the right panel in Figure A.4, performs poorly due to its incapability of timing spikes and drops. Similarly, the reserve margin prognosis also produces results below average. Like daily wind power generation, hourly measurements and forecasts of the wind power generation generate poor timing of spikes and many of the transition probabilities are noise that seems incompatible with the Nordic market, at least as a single explanatory variable. The price dynamics of the German market are arguably better explained by wind power production than the Nordic market, since its trend is less seasonal, the noise that it carries resembles the spread in EEX more than in NPS, and spikes and drops in EEX are more equally distributed over a whole year.

Finally, we conclude that consumption and production improve the timing of spikes and perform best overall. While consumption forecasts yield similar results, the production counterpart is not as precise. This finding can arguably be traced to the relatively small sample size. By contrast, wind power generation does not time spikes as well, which is probably due to its lack of seasonal pattern that is prominent in the log spread, cf. Botterud et al. (2010). The shape of the hourly transition probabilities is less satisfactory overall. A simulation study shall be conducted in order to analyze this further. Better timing could probably be accomplished by merging the variables into one single model or by for instance examining appropriate ratios or time lags of the exogenous processes.

It would also be interesting to analyze area prices rather than system prices in e.g. the Nordic market in order to better understand regional variations. Wind power generation would in particular be interesting to analyze on a country- or even area-specific level.

### 5.3.2 Price Simulations

Prices are simulated after adding the deterministic seasonal components to the calibrated models and inverting the transformed prices. One hundred Markov chains are first simulated, whereupon mean prices are obtained in the Monte Carlo sense. The simulations are delimited to the optimal explanatory variables, viz. consumption, consumption forecasts and production.

The simulations are shown in Figure 5.7 for daily prices and Figure 5.8 for hourly prices. The overall fit is good, but it lacks the magnitude of the spikes observed in the market. The main reason behind this can be explained by the fact that the Monte Carlo simulations generate mean prices. The choice of spike distributions also affects the simulations, that is log-normal spikes capture extreme prices better, though this effect is secondary. The root mean square errors for daily consumption, production and consumption prognosis are 6.30, 6.68 and 6.75, respectively, hence indicating that consumption is the best explanatory variable among the three. Similar results are obtained on an hourly basis, where even greater deviations in prices can be observed. Even though gamma spikes are not able to generate such outliers, the overall fit is yet again satisfactory.

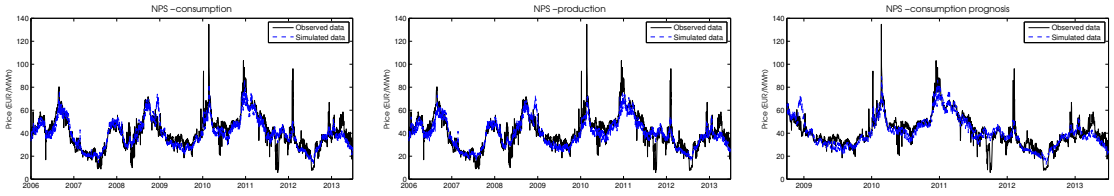



Figure 5.7: Monte Carlo simulations of daily prices in NPS using three-state MRS models calibrated with Vašíček dynamics, gamma spikes and exogenous processes. While the overall fit is satisfactory, spikes and drops are naturally not attaining the extreme magnitudes observed in the market.

Both daily and hourly prediction errors of the Monte Carlo simulations are displayed in Figure 5.9. From the figure it is clear that the models are able to predict the majority of the prices relatively well. It is also clear that the magnitudes of spikes are harder to estimate than drops. This inaccuracy is, however, evident irrespective of the choice of exogenous process.

While the timing of spikes and drops has been greatly improved, the simulation studies indicate some shortcomings of the calibration results. Inclusion of more exogenous processes could improve the timing of spikes, especially off-cycle spikes. The choice of spike distribution is, as already mentioned, not final and further simulations studies comparing such distributions — both in-sample and out-of-sample — are needed. Examination of

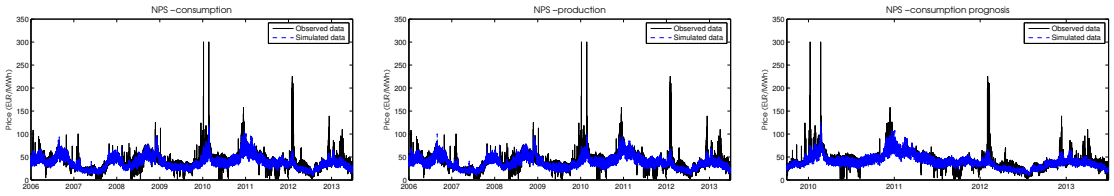



Figure 5.8: Monte Carlo simulations of hourly prices in NPS using three-state MRS models calibrated with Vašíček dynamics, gamma spikes and exogenous processes. While the overall fit is satisfactory, spikes and drops are naturally not attaining the extreme magnitudes observed in the market.

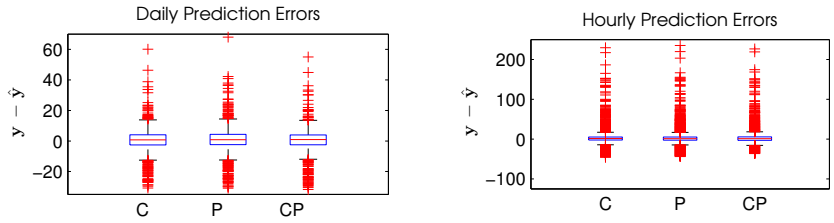



Figure 5.9: Prediction errors of Monte Carlo simulations of daily and hourly prices in NPS using three-state MRS models calibrated with Vašíček dynamics, gamma spikes and exogenous processes.

calibrations of several inhomogeneous models with different spike distributions would in particular be worthwhile.

## 5.4 Market Comparison

After identifying spikes it is interesting to make use of the calibration results to further understand market integration. From a risk management perspective it is crucial to understand the relationship between markets, in order to eliminate or at least reduce cross-market risk.

### 5.4.1 Extreme Event Analysis

In order to examine cospiking between markets, we implement the ad-hoc method introduced by Lindström and Regland (2012) which measures the unconditional probability of a latent state being an extreme event $E \subset \mathsf{X}$ as the Monte Carlo estimate of the maximum a posteriori estimates such that

$$\mathbb{P}(E) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_E(\hat{X}_t).$$

Consequently, the conditional probability is obtained as

$$\mathbb{P}(E_i \mid E_j) = \frac{\mathbb{P}(E_i, E_j)}{\mathbb{P}(E_i)} = \frac{\sum_{t=1}^{T} \mathbb{1}_E(\hat{X}_{i,t})\mathbb{1}_E(\hat{X}_{j,t})}{\sum_{t=1}^{T} \mathbb{1}_E(\hat{X}_{i,t})},$$

i.e the conditional probability $\mathbb{P}(E_{EEX} = S \mid E_{NPS} = S)$ signifies the probability of EEX experiencing spikes at the same time as NPS. Two stochastic processes shall suffice for the forthcoming analysis, though this measure can easily be extended to arbitrarily many state processes. Lindström and Regland (2012) advocate this asymmetric measure because of its main advantage over symmetric measures such as correlation, which provide less information.

Cospiking is investigated on both a daily and hourly basis, see Table 5.5, which presents the conditional probability of a market in a row experiencing spikes or drops at the same time as another market in the corresponding column. The conditional probabilities on a daily basis are quite low, but still significantly higher than the ones obtained by Lindström and Regland (2012). Above all, the German market seems to be more reliant on the Nordic market than vice versa, as it cospikes roughly twice as many times compared with NPS. The correlation between EEX and NPS is, besides their dissimilar allocation of energy sources and their geographical propinquity, probably mainly due to relatively small integration.

Other markets have been found to experience higher correlation with the German market, see e.g. de Jong (2006) and Lindström and Regland (2012). As de Jong (2006) points out, this can be related to the large share of hydropower produced in the Nordic countries. In comparison with bordering countries of Germany, the relatively long distance between the German market and hydroelectric power stations in the Nordic countries impedes inflow to and outflow from the Nordic market and implies higher exchange between closer markets. The more stable hydropower therefore only affects the Nordic area, whereas the price sensitivity of the German market remains unchanged.

Table 5.5: Conditional probabilities of a market in a row experiencing spikes or drops at the same time as a market in the corresponding column from 1 January 2006 to 30 June 2013. Peak hours are defined as 07:00 – 22:00 CET.

| | Spikes | | Drops | |
|---|---|---|---|---|
| Market | EEX | NPS | EEX | NPS |
| EEX daily | 1.0000 | 0.2323 | 1.0000 | 0.2731 |
| NPS daily | 0.0985 | 1.0000 | 0.1279 | 1.0000 |
| EEX hourly | 1.0000 | 0.2630 | 1.0000 | 0.2472 |
| NPS hourly | 0.4370 | 1.0000 | 0.2779 | 1.0000 |
| EEX peak | 1.0000 | 0.0775 | 1.0000 | 0.2226 |
| NPS peak | 0.1706 | 1.0000 | 0.2525 | 1.0000 |
| EEX off-peak | 1.0000 | 0.4972 | 1.0000 | 0.2890 |
| NPS off-peak | 0.6039 | 1.0000 | 0.3092 | 1.0000 |

More interestingly, while EEX experiences daily and hourly spikes and drops at approximately the same amount of times, NPS clearly shows a closer relation to simulta-

neous intraday spikes in EEX than vice versa. The cospiking between the Nordic and German markets is in that case more than four times than what is observed on a daily basis, whereas the conditional probability of drops is more than twice as high. A great part of this can be explained by the vast amount of cospikes occurring during off-peak hours. While the breakdown of intraday spikes discloses that drops occur at roughly the same proportion in both markets during peak and off-peak hours, NPS experiences a larger amount of up-spikes during peak and off-peak hours at the same time as EEX than vice versa.

In summary, the Nordic market is less dependent on EEX on a daily level than vice versa, whereas an inverse relationship is evident intraday. Not only does the decomposition of intraday spikes evince that most cospikes appear during off-peak hours, but it also shows that NPS has a twofold experience of up-spikes during peak hours compared with the German market.

Chapter 6

# Conclusion

## 6.1 Conclusive Summary

This thesis provides a guided tour of Markov-switching models and how to apply them to electricity spot prices as well as viable methods of inference. Both maximum likelihood and the EM algorithm have been described and discussed. Closed-form expressions for most parameters have also been presented in the latter case. Simulation studies have illustrated the convergence of both techniques and compared them with each other. The methods yielded relatively small errors, even for inhomogeneous transitions, and were considered equally accurate.

Stylized facts of electricity prices have been implemented in order to improve the modelling of day-ahead spot prices on a daily and hourly basis. The occurrence of power spikes has been emphasized and modelled with non-linear two- and three-state MRS models, where the non-linearity has been embedded in both state and space, in order to capture bidirectional spikes. Prices transformations and deseasonalization techniques have been discussed and implemented before proceeding to model calibration. The ubiquitous logarithmic transformation has received a lot of attention since negative prices have been observed in some electricity markets due to regulatory changes. The problems that arise from such prices undermine the use of the standard log transformation. Although a remedy of the form (4.1) has been proposed in this thesis in order to address this particular issue, the inverse hyperbolic sine function has been used due to its previous successes, see e.g. Schneider (2012). The IR setting has been adopted in order to allow for rapid changes between price levels. The parsimonious method proposed by Janczura and Weron (2012) has been implemented to compute the latent lagged values of the autoregressive submodels.

The study was confined to the German and Nordic markets because of their size and liquidity as well as their popularity among researchers. The discussion has in particular revolved around exogenous data in order to encircle a set of interesting explanatory variables, from which a feasible subset finally was chosen, namely consumption, production, reserve margin and wind power generation as well as forecasts thereof for the Nordic market, whereas gas and EUAs were used for EEX.

Standard models have been used as benchmarks in order to compare them with more complex models. Calibration results have shown that two-state models are not enough to encompass all types of spikes. An additional state was therefore introduced and several tests concluded that the refined models outperformed the two-state models. The optimal models for the German and Nordic markets are three-state models with log-normal and gamma spikes, respectively. After having placed CKLS dynamics under scrutiny and tested them meticulously, we also reached the conclusion that Vašíček dynamics fit the majority of the prices on both daily and hourly basis.

Above all, the extension to non-linear state equations has showed that consumption and production as well as forecasts thereof are able to better capture price dynamics. The models have identified typical seasonal cycles such that spikes (drops) are more likely during winter (summer) seasons, which is in line with our intuition and what can be observed in the market. While the magnitude of extreme prices has been encompassed by the spike distributions, this pivotal extension has led to better timing of spikes — a feature that has been requested by the literature for quite some time. Settled and predicted consumption yielded the best overall results and were able to drive prices according to seasonal cycles.

A market comparison between EEX and NPS has also been carried out in the spirit of Lindström and Regland (2012) after calibrating the models. The probability of experiencing simultaneous spikes has been studied in this context. Like earlier research has shown, the German market was found to be more reliant on the Nordic market on a daily basis than vice versa. Intraday spikes in the Nordic market were, by contrast, more dependent on EEX. Interestingly, a large number of the cospikes stemmed from off-peak hours.

## 6.2 Future Research Outlook

This thesis has shown that electricity prices, indeed, depend on intimately related factors. Yet there remain explanatory variables of interest that have not been studied in this thesis, where the most obvious one probably is temperature. In contrast to the exogenous data used in this thesis, neither are hourly data always available nor is it obvious how some data should be treated to fit the model framework. Even though for instance ambient temperature could be incorporated into any MRS model, it seems inappropriate to do so for at least some markets that cover huge territories with large variations in temperature, cf. the Nordic countries. In order to circumvent this problem it would probably be better to confine the analysis to separate countries or areas. Similarly, wind power generation could arguably describe area prices better than system prices.

A hinge for conducting further analyses is the limited availability of data. More countries and areas as well as more explanatory variables are of interest in order to understand the similarities and differences between regions. Larger sample sizes would also reduce the parameter variance so that more complex models could be built by combining several exogenous processes with different weights that govern state transitions. A wider range of frequencies could in this way improve the timing of off-cycle spikes.

Slight adjustments of spikes are also preferable in order to fit prices better. One suggestion is to examine different combinations of families of distributions separately for spikes and drops. Further fine-tuning is preferable in order to improve in-sample fit and out-of-sample forecasts of daily and particularly hourly prices. Another appealing idea is to leave spikes unspecified as in Eichler and Türk (2013). Moreover, the problem of transforming negative prices is also not settled, even though transformations such as the inverse hyperbolic sine function or the shifted log function (4.1) resemble the standards log function with respect to shape and asymptotic property.

One of the conundrums that is left for further investigation is the deseasonalization. Although Nowotarski et al. (2013) make a great attempt when examining a wide range of models, a more definite technique would be advantageous in order to ease comparisons of models. The deseasonalization could be made more efficient, perhaps by adapting wavelets or similar filtering techniques. A more adjacent alternative to monthly forward prices is to filter out the nearest forward-looking spike information by transforming the forward prices and reducing the week-ahead information, for instance by using weekly forwards.

Finally, even though the model framework has been applied to electricity markets, it is straightforward to adapt the framework to other fields of interest. The non-linearity of the models not only compliments the need to explicitly specify the variance process per se, but it also allows for indirect relationships between multiple processes to be studied, since an exogenous process first and foremost affects the underlying Markov chain. This thesis has provided several examples of how such relationships can be studied in electricity markets.

Appendix A

# Calibration Results for Exogenous Data

## A.1 European Energy Exchange

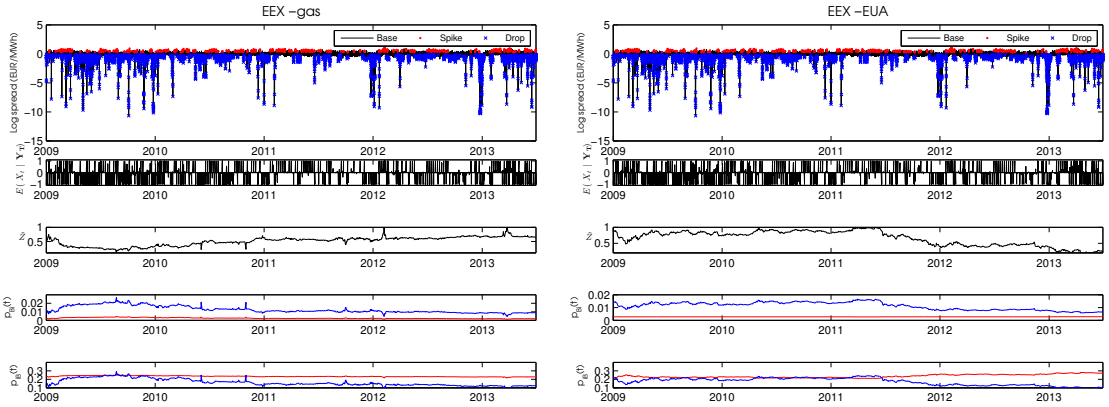



Figure A.1: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and log-normal spikes for hourly day-ahead prices in EEX using EUAs and the mean of regional gas prices. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

## A.2 Nord Pool Spot

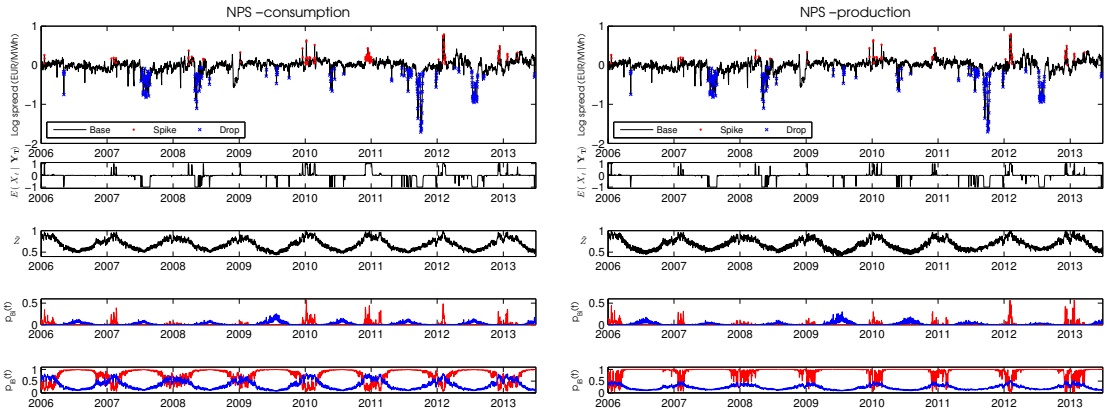



Figure A.2: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for daily day-ahead prices in NPS using consumption and production. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

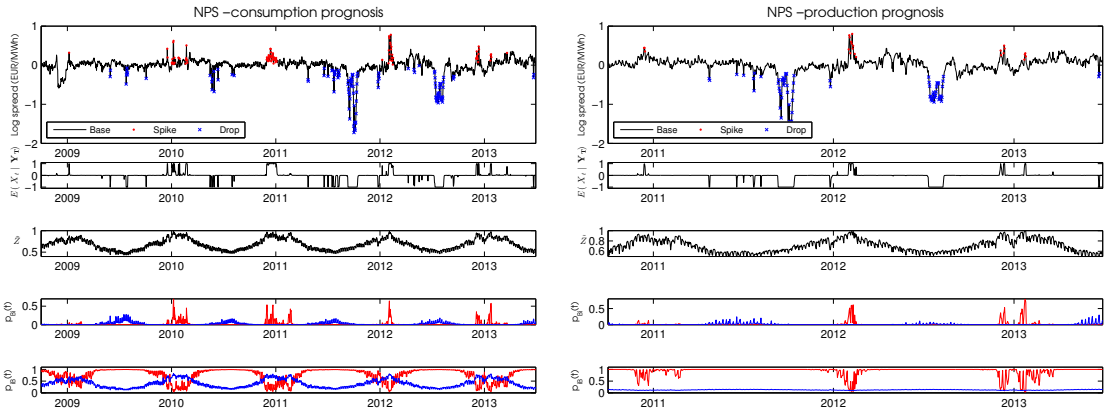



Figure A.3: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for daily day-ahead prices in NPS using consumption and production forecasts. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

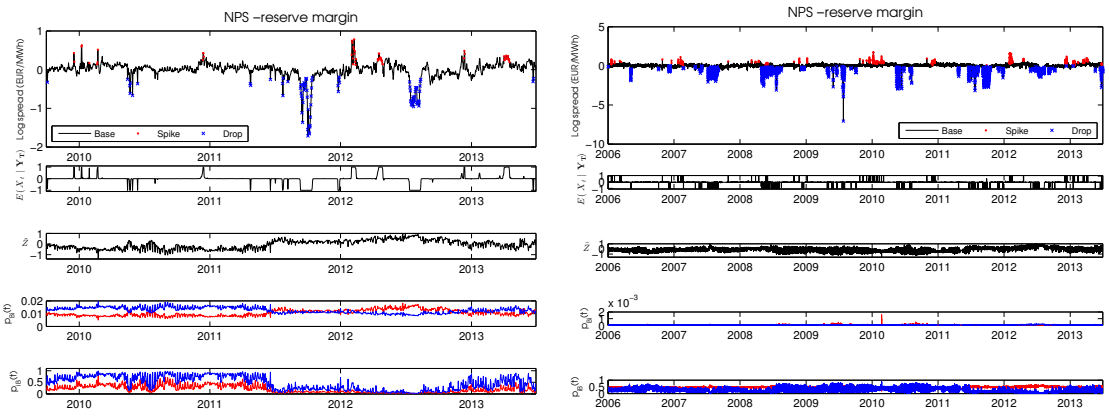



Figure A.4: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for daily (left panel) and hourly (right panel) day-ahead prices in NPS using the reserve margin. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

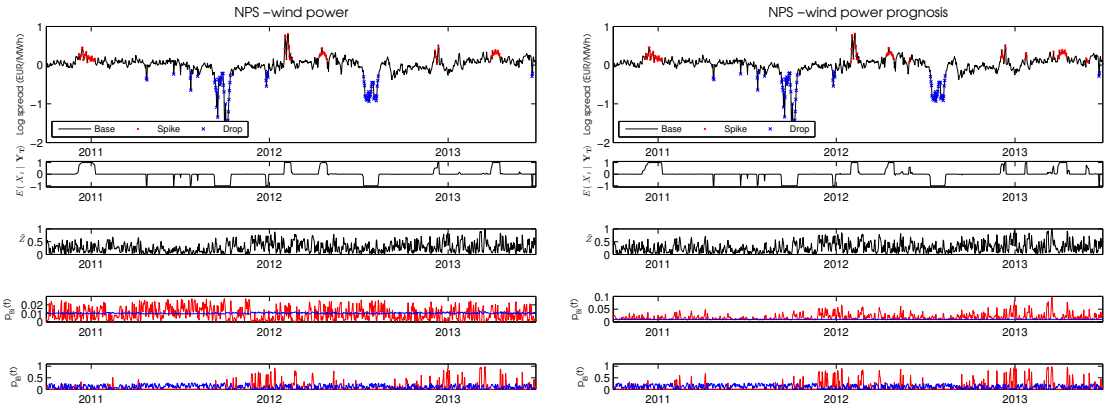



Figure A.5: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for daily day-ahead prices in NPS using wind power generation and forecasts thereof. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

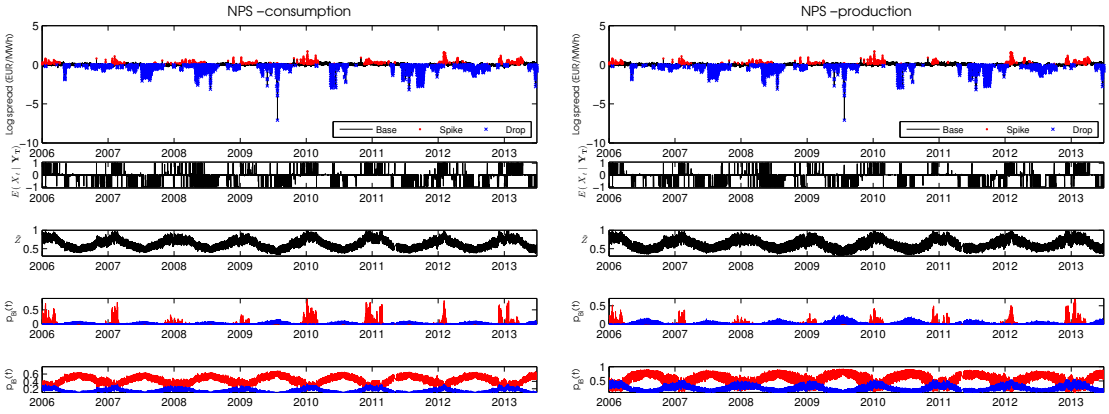



Figure A.6: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for hourly day-ahead prices in NPS using consumption and production. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

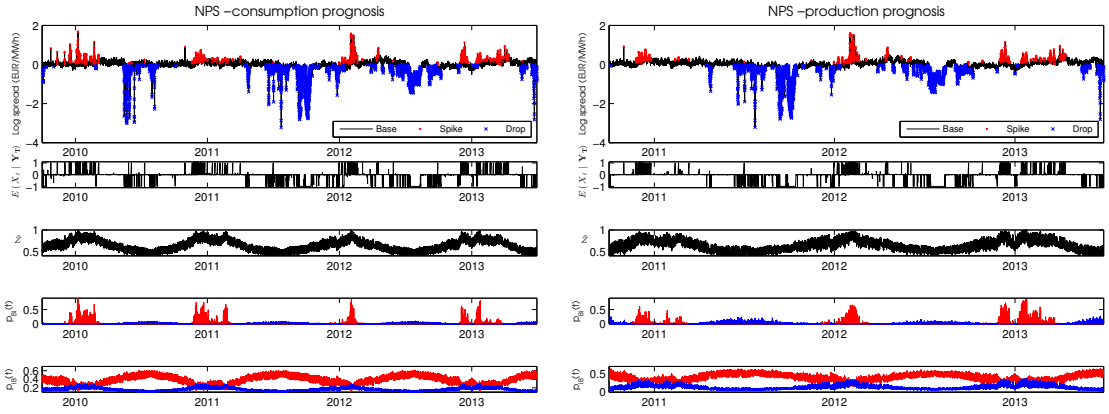



Figure A.7: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for hourly day-ahead prices in NPS using consumption and production forecasts. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

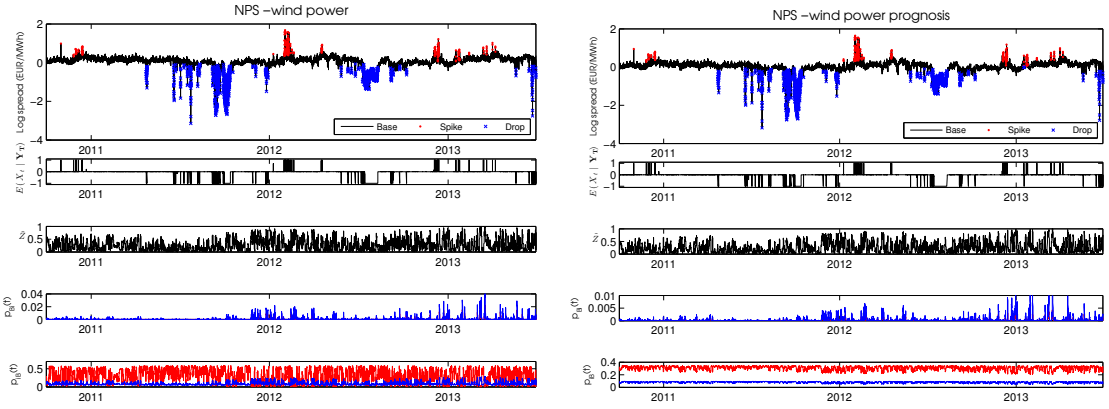



Figure A.8: Calibration of inhomogeneous three-state MRS models with Vašíček dynamics and gamma spikes for hourly day-ahead prices in NPS using wind power generation and forecasts thereof. The two uppermost rows illustrate the identification of spikes, whereas the middle row shows the normalized exogenous process. The two lowermost rows highlight the evolution of the transition probabilities.

# Bibliography


Anderson, C. Lindsay, and Matt Davison. "A Hybrid System-Econometric Model for Electricity Spot Prices: Considering Spike Sensitivity to Forced Outage Distributions." *IEEE Transactions on Power Systems* Vol. 3, No. 4, pp. 573–592, 2008.

Bachelier, Louis. "Théorie de la Spéculation." *Annales Scientifiques de l'École Normale Supérieure* Vol. 23, No. 3, pp. 927–937, 1900.

Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *The Annals of Mathematical Statistics* Vol. 41, No. 1, pp. 164–171, 1970.

Becker, Ralf, Stan Hurn, and Vlad Pavlov. "Modelling Spikes in Electricity Prices." *Economic Record* Vol. 83, No. 263, pp. 371–382, 2007.

Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* Vol. 81, No. 3, pp. 637–654, 1973.

Black, Fisher, and Piotr Karasinski. "Bond and Option Pricing when Short Rates are Lognormal." *Financial Analyst Journal* Vol. 47, No. 4, pp. 52–59, 1991.

Blanco, Carlos, David Soronow, and Paul Stefiszyn. "Multi-Factor Models of the Forward Price Curve." *Commodities Now* Vol. 2, No. 3, pp. 80–83, 2002.

Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* Vol. 31, No. 3, pp. 307–327, 1986.

Botterud, Audun, Tarjei Kristiansen, and Marija D. Ilic. "The Relationship Between Spot and Futures Prices in the Nord Pool Electricity Market." *Energy Economics* Vol. 32, No. 5, pp. 967–978, 2010.

Brennan, Michael J., and Eduardo S. Schwartz. "Analyzing Convertible Bonds." *Journal of Financial and Quantitative Analysis* Vol. 15, No. 4, pp. 907–929, 1980.

Brzeźniak, Zdzisław, and Tomasz Zastawniak. *Basic Stochastic Processes: A Course Through Exercises*. Springer-Verlag, London Berlin Heidelberg, 2002.

Calvet, Laurent E., and Adlai J. Fisher. "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes." *Journal of Financial Econometrics* Vol. 2, No. 1, pp. 49–83, 2004.

Cappé, Olivier, Eric Moulines, and Tobias Rydén. *Inference in Hidden Markov Models*. Springer, New York, 2005.

Cartea, Álvaro, Marcelo G. Figueroa, and Hélyette Geman. "Modelling Electricity Prices with Forward Looking Capacity Constraints." *Applied Mathematical Finance* Vol. 16, No. 2, pp. 103–122, 2009.

Chan, K. C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders. "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate." *Journal of Finance* Vol. 47, No. 3, pp. 1209–1227, 1992.

Chang, Kuang-Liang. "Volatility Regimes, Asymmetric Basis Effects and Forecasting Performance: An Empirical Investigation of the WTI Crude Oil Futures Market." *Energy Economics* Vol. 34, No. 1, pp. 294–306, 2012.

Christensen, Tim M., A. Stan Hurn, and Kenneth A. Lindsay. "Forecasting Spikes in Electricity Prices." *International Journal of Forecasting* Vol. 28, No. 2, pp. 400–411, 2012.

Cox, John C. *Notes on Option Pricing: I: Constant Elasticity of Diffusions*. Stanford University, 1975.

Cox, John C., Jonathan E. Ingersoll, and Jr. Stephen A. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica* Vol. 53, No. 2, pp. 385–408, 1985.

Cruz, Alberto, Antonio Muñoz, Juan Luis Zamora, and Rosa Espínola. "The Effect of Wind Generation and Weekday on Spanish Electricity Spot Price Forecasting." *Electric Power Systems Research* Vol. 81, No. 10, pp. 1924–1935, 2011.

Csiszár, Imre, and Paul C. Shields. "The Consistency of the BIC Markov Order Estimator." *The Annals of Statistics* Vol. 28, No. 6, pp. 1601–1619, 2000.

Davison, Matt, C. Lindsay Anderson, Ben Marcus, and Karen Anderson. "Development of a Hybrid Model for Electrical Power Spot Prices." *IEEE Transactions on Power Systems* Vol. 17, No. 2, pp. 257–264, 2002.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* Vol. 39, No. 1, pp. 1–38, 1977.

Deng, Shijie. "Stochastic Models of Energy Commodity Prices and Their Applications: Mean-Reversion with Jumps and Spikes." Working Paper PWP-073, Program on Workable Energy Regulation, University of California, Berkley, 1998.

Diebold, Francis X., Joon-Haeng Lee, and Gretchen C. Weinbach. "Regime Switching with Time-Varying Transition Probabilities" in *Nonstationary Time Series Analysis and Cointegration*, edited by Colin P. Hargreaves. Oxford University Press, Oxford, 1994.

Eichler, Michael, and Dennis Türk. "Fitting Semiparametric Markov Regime-Switching Models to Electricity Spot Prices." *Energy Economics* Vol. 36, pp. 614–624, 2013.

Engle, Robert F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* Vol. 50, No. 4, pp. 987–1007, 1982.

Erdős, Paul, William Feller, and Harry Pollard. "A Property of Power Series with Positive Coefficients." *Bulletin of the American Mathematical Society* Vol. 55, No. 2, pp. 201–204, 1949.

Erlwein, Christina, Fred Espen Benth, and Rogemar Mamon. "HMM Filtering and Parameter Estimation of an Electricity Spot Price Model." *Energy Economics* Vol. 32, No. 5, pp. 1034–1043, 2010.

Ethier, Robert, and Timothy D. Mount. "Estimating the Volatility of Spot Prices in Restructured Electricity Markets and the Implications for Option Values." PSerc Working Paper, 98–31, 1998.

Eydeland, Alexander, and Krzysztof Wolyniec. *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging*. John Wiley & Sons, Hoboken, 2003.

Fabra, Natalia, and Juan Toro. "Price Wars and Collusion in the Spanish Electricity Market." *International Journal of Industrial Organization* Vol. 23, Nos. 3–4, pp. 155–181, 2005.

Filardo, Andrew J. "Business-Cycle Phases and Their Transitional Dynamics." *Journal of Business & Economic Statistics* Vol. 12, No. 3, pp. 299–308, 1994.

Filardo, Andrew J., and Stephen F. Gordon. "Business Cycle Durations." *Journal of Econometrics* Vol. 85, No. 1, pp. 99–123, 1998.

Frühwirth-Schnatter, Sylvia. *Finite Mixture and Markov Switching Models*. Springer, New York, 2006.

Glasserman, Paul. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2003.

Goldfeld, Stephen M., and Richard E. Quandt. "A Markov Model for Switching Regressions." *Journal of Econometrics* Vol. 1, No. 1, pp. 3–16, 1973.

Haldrup, Niels, and Morten Ørregaard Nielsen. "A Regime Switching Long Memory Model for Electricity Prices." *Journal of Econometrics* Vol. 135, Nos. 1–2, pp. 349–376, 2006.

Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* Vol. 57, No. 2, pp. 357–384, 1989.

Hamilton, James D. "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics* Vol. 45, Nos. 1–2, pp. 39–70, 1990.

Hamilton, James D. *Time Series Analysis*. Princeton University Press, Princeton, 1994.

Hamilton, James D., and Baldev Raj. "New Directions in Business Cycle Research and Financial Analysis." *Empirical Economics* Vol. 27, No. 2, pp. 149–162, 2002.

Huisman, Ronald. "The Influence of Temperature on Spike Probability in Day-Ahead Power Prices." *Energy Economics* Vol. 30, No. 5, pp. 2697–2704, 2008.

Huisman, Ronald, Christian Huurman, and Ronald Mahieu. "Hourly Electricity Prices in Day-Ahead Markets." *Energy Economics* Vol. 32, No. 2, pp. 240–248, 2007.

Huisman, Ronald, and Ronald Mahieu. "Regime Jumps in Electricity Prices." *Energy Economics* Vol. 25, No. 5, pp. 425–434, 2003.

Hull, John, and Alan White. "Pricing Interest-Rate-Derivative Securities." *Review of Financial Studies* Vol. 3, No. 4, pp. 573–592, 1990.

Hürzeler, Markus, and Hans R. Künsch. "Approximating and Maximising the Likelihood for a General State-Space Model" in *Sequential Monte Caro Methods in Practice*, edited by Arnaud Doucet, Nando de Freitas, and Neil Gordon. Springer-Verlag, New York, 2001.

Janczura, Joanna, Stefan Trück, Rafał Weron, and Rodney C. Wolff. "Identifying Spikes and Seasonal Components in Electricity Spot Price Data: A Guide to Robust Modeling." *Energy Economics* Vol. 38, pp. 96–110, 2013.

Janczura, Joanna, and Rafał Weron. "An Empirical Comparison of Alternate Regime-Switching Models for Electricity Spot Prices." *Energy Economics* Vol. 32, No. 5, pp. 1059–1073, 2010.

Janczura, Joanna, and Rafał Weron. "Efficient Estimation of Markov Regime-Switching Models: An Application to Electricity Spot Prices." *AStA Advances in Statistical Analysis* Vol. 96, No. 3, pp. 385–407, 2012.

de Jong, Cyriel. "The Nature of Power Spikes: A Regime-Switching Approach." *Studies in Nonlinear Dynamics & Econometrics* Vol. 10, No. 3, Article 3, 2006.

de Jong, Cyriel, and Ronald Huisman. "Option Formulas for Mean-Reverting Power Prices with Spikes." ERIM Report Series Research in Management, No. ERS-2002-96-F&A, Erasmus University, Rotterdam, 2002.

de Jong, Cyriel, and Stefan Schneider. "Cointegration Between Gas and Power Spot Prices." *Journal of Energy Markets* Vol. 2, No. 3, pp. 27–46, 2009.

Kanamura, Takashi, and Kazuhiko Ōhashi. "On Transition Probabilities of Regime Switching in Electricity Prices." *Energy Economics* Vol. 30, No. 3, pp. 1158–1172, 2008.

Karakatsani, Nektaria V., and Derek W. Bunn. "Forecasting Electricity Prices: The Impact of Fundamentals and Time-Varying Coefficients." *International Journal of Forecasting* Vol. 24, No. 4, pp. 764–785, 2008.

Keles, Dogan, Massimo Genoese, Dominik Möst, and Wolf Fichtner. "Comparison of Extended Mean-Reversion and Time Series Models for Electricity Spot Price Simulation Considering Negative Prices." *Energy Economics* Vol. 34, No. 4, pp. 1012–1032, 2012.

Kim, Chang-Jin. "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics* Vol. 60, Nos. 1–2, pp. 1–22, 1994.

Krogh, Anders, Michael Brown, I. Saira Mian, Kimmen Sjölander, and David Haussler. "Hidden Markov Models in Computational Biology: Applications to Protein Modeling." *Journal of Molecular Biology* Vol. 235, No. 5, pp. 1501–1531, 1994.

Lindström, Erik, and Fredrik Regland. "Modeling Extreme Dependence Between European Electricity Markets." *Energy Economics* Vol. 34, No. 4, pp. 899–904, 2012.

Lu, Zhan-Qian, and L. Mark Berliner. "Markov Switching Time Series Models with Application to a Daily Runoff Series." *Water Resources Research* Vol. 35, No. 2, pp. 523–534, 1999.

Lucia, Julio J., and Eduardo S. Schwartz. "Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange." *Review of Derivatives Research* Vol. 5, No. 1, pp. 5–50, 2002.

Løland, Anders, Egil Ferkingstad, and Mathilde Wilhelmsen. "Forecasting Transmission Congestion." *Journal of Energy Markets* Vol. 5, No. 3, pp. 65–83, 2012.

Martinez Peria, Maria Soledad. "A Regime-Switching Approach to the Study of Speculative Attacks: A Focus on EMS Crises." *Empirical Economics* Vol. 27, No. 2, pp. 299–334, 2002.

Metropolis, Nicholas, and Stanisław Ulam. "The Monte Carlo Method." *Journal of the American Statistical Association* Vol. 44, No. 247, pp. 335–341, 1949.

Misiorek, Adam, Stefan Trück, and Rafał Weron. "Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models." *Studies in Nonlinear Dynamics & Econometrics* Vol. 10, No. 3, Article 2, 2006.

Mount, Timothy D., Yumei Ning, and Xiaobin Cai. "Predicting Price Spikes in Electricity Markets Using a Regime-Switching Model with Time-Varying Parameters." *Energy Economics* Vol. 28, No. 1, pp. 62–80, 2006.

Musiela, Marek, and Marek Rutkowski. *Martingale Methods in Financial Modelling*, 2nd edition. Springer-Verlag, Berlin Heidelberg, 2005.

Norris, James R. *Markov Chains*. Cambridge University Press, Cambridge, 1998.

Nowotarski, Jakub, Jakub Tomczyk, and Rafał Weron. "Robust Estimation and Forecasting of the Long-Term Seasonal Component of Electricity Spot Prices." *Energy Economics* Vol. 39, pp. 13–27, 2013.

Rabiner, Lawrence R. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* Vol. 77, No. 2, pp. 257–286, 1989.

Rambharat, B. Ricky, Anthony E. Brockwell, and Duane J. Seppi. "A Threshold Autoregressive Model for Wholesale Electricity Prices." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* Vol. 54, No. 2, pp. 287–299, 2005.

Regland, Fredrik, and Erik Lindström. "Independent Spike Models: Estimation and Validation." *Czech Journal of Economics and Finance* Vol. 62, No. 2, pp. 180–196, 2010.

Robert, Christian P., and George Casella. *Monte Carlo Statistical Methods*, 2nd edition. Springer, New York, 2004.

Robinson, Terry A. "Electricity Pool Prices: A Case Study in Nonlinear Time-Series Modelling." *Applied Economics* Vol. 32, No. 5, pp. 527–532, 2000.

Rudin, Walter. *Principles of Mathematical Analysis*, 3rd edition. MacGraw-Hill, 1976.

Schneider, Stefan. "Power Spot Price Models with Negative Prices." *Journal of Energy Markets* Vol. 4, No. 4, pp. 77–102, 2012.

Schwarz, Gideon. "Estimating the Dimension of a Model." *The Annals of Statistics* Vol. 6, No. 2, pp. 461–464, 1978.

Szkuta, B.R., L.A. Sanabria, and T.S. Dillon. "Electricity Price Short-Term Forecasting Using Artificial Neural Networks." *IEEE Transactions on Power Systems* Vol. 1, No. 3, pp. 851–857, 1999.

Uhlenbeck, George E., and Leonard S. Ornstein. "On the Theory of the Brownian Motion." *Physical Review* Vol. 36, No. 5, pp. 823–841, 1930.

Vašíček, Oldřich. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics* Vol. 5, No. 2, pp. 177–188, 1977.

Wang, A.J., and B. Ramsay. "A Neural Network Based Estimator for Electricity Spot-Pricing with Particular Reference to Weekend and Public Holidays." *Neurocomputing* Vol. 23, Nos. 1–3, pp. 47–57, 1998.

Weron, Rafał. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. John Wiley & Sons, Chichester, 2006.

Weron, Rafał. "Heavy-Tails and Regime-Switching in Electricity Prices." *Mathematical Methods of Operations Research* Vol. 69, No. 3, pp. 457–473, 2009.

Weron, Rafał, Michael Bierbrauer, and Stefan Trück. "Modeling Electricity Prices: Jump Diffusion and Regime Switching." *Physica A: Statistical Mechanics and Its Applications* Vol. 336, Nos. 1–2, pp. 39–48, 2004.

Wong, Chun Shan, and Wai Keung Li. "On a Logistic Mixture Autoregressive Model." *Biometrika* Vol. 88, No. 3, pp. 833–846, 2001.

Zachmann, Georg. "A Stochastic Fuel Switching Model for Electricity Prices." *Energy Economics* Vol. 35, pp. 5–13, 2013.

Zucchini, Walter, and Iain L. MacDonald. *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman & Hall, Boca Raton, 2009.

Master's Theses in Mathematical Sciences 2013:E66
ISSN 1404-6342

LUTFMS-3233-2013

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/