Numerical Methods for Stochastic Differential Equations with Applications to Electricity Futures

Fall 2022
October 20, 2023

Nicolaj Hans Nielsen, s184335
Anton Ruby Larsen, s174356

Danmarks
Tekniske Universitet

Contents

A | Introduction

In this report, we will develop a method to quantify the market risk premium in electricity markets and apply stochastic Runge-Kutta methods to improve our estimation of parameters.
The market risk premium is the additional premium that market players in commodity markets are willing to pay to hedge the price of future volumes. This market premium is important for the market players as it quantifies the risk involved in energy trading, it plays a great role in transfer pricing, and most importantly it can be used to ensure more adequate derivative pricing methods. Adequate pricing and functional energy markets are essential for generators, retailers, and governments as they are the main factors in the energy trilemma of obtaining *secure, sustainable, and affordable* energy [1]. We will consider the premium as the difference between the expected spot prices and the observed price of the futures contracts which is based on ideas introduced in [2] and [3]. To do this, we will construct a model for the spot price, which is composed of 3 different parts, $S_t = \Lambda(t) + X(t) + Y(t)$. Here $\Lambda(t)$ is a simple deterministic component that contains the general price relationships that characterise the market. $X(t)$ will be the diffusive component to describe the non-extreme variation and mean reversion of the spot price. $Y(t)$ is a Lévy-process to capture the intrinsic jumps in electricity which is a direct effect of a spiky demand profile and the inability to effectively store electricity. Then we will cover the link between the spot and futures price, describe the methods to make a change of measure, and introduce how we will fit the spot model. To estimate the parameters of $X(t)$, we introduce the concept of stochastic Runge-Kutta methods which are both more precise and more computationally efficient than traditional methods. We will fit the spot price model on EPEX spot prices in the period from 2002 to 2005 and calculate the risk premium on the baseload EEX futures in the period from 2002 to 2005.

B | Pricing of Energy Futures

In this section, we will cover how we can link the electricity spot price to the futures prices. In fig. B.1, we see the spot price we consider and one of the available futures series. In our modelling setup, the general idea is to construct a model for the spot price based on a deterministic part, a diffusive part, and a levy part and link it to the future price using the expected value and a change of measure. In the end, we will get to an expression like,

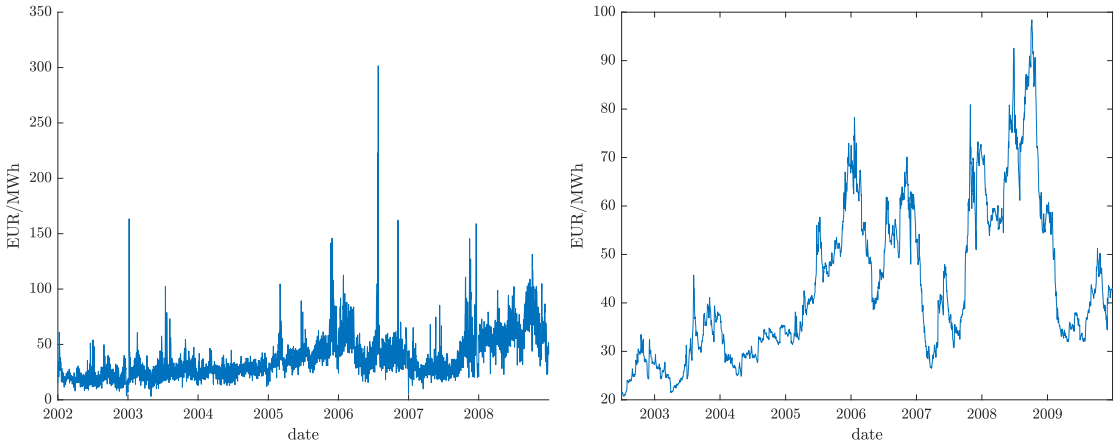
$$F^{\mathbb{Q}}_{\mathcal{F}_t}(t, T_1, T_2) = \mathbb{E}^{\mathbb{Q}}\left[\int_{T_1}^{T_2} \frac{1}{T_2 - T_1} S_u \, \mathrm{d}u \middle| \mathcal{F}_t\right], \tag{B.0.1}$$

where $\mathcal{F}_t$ is the historical filtration of the spot price, $\mathbb{Q}$ is a risk-neutral measure, and $[T_1, T_2]$ is the interval, which the contract covers e.g. only January 2023, the second quarter 2023 or a yearly contract covering all of 2023. In our case, we will consider monthly contracts only.
In the following, we will cover the necessary concepts to understand the components of eq. (B.0.1) in the following order:

1. Change of probability measure and the Girsanov theorem which makes us able to construct a risk-neutral pricing measure.

2. Introduction of the sport price model and the properties of the constituent components.

3. We will derive how the future price can be calculated using the expected future value of the spot price model under a change of probability measure.



(a) EPEX spot price in the period Considered. **(b)** M1 EEX contract in the period considered.

Figure B.1 – The EPEX spot price over the period and available M1 futures data.

B.1 Change of Probability Measure and Risk-neutral Pricing

In the financial literature, the Girsanov transformation and Radon Nikodym derivative are central for the pricing of derivatives as they give a way to transform a stochastic differential equation from the real-world measure to a so-called risk-neutral measure. The risk-neutral measure is to be understood as an equivalent martingale measure that can impose the Markov property on a stochastic process by changing the probability measure. In the following, we will introduce the Girsanov transformation through the example of a geometric Brownian motion.

B.1.1 Girsanov Transformation for the Geometric Brownian Motion

Let the price of the commodity be defined as an Itô stochastic process, $\{S_t : t \geq 0\}$, defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ with $\{\mathcal{F}_t : t \geq 0\}$ governed by the SDE commonly referred to as the geometric Brownian motion,

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \, \mathrm{d}B_t \qquad S_0 = s. \tag{B.1.1}$$

Where $\mu \in \mathbb{R}$ is the drift parameter, and with some noise intensity σS_t with $\sigma \in \mathbb{R}_+$. The economic interpretation of this model for a stock is:

- μ: is the expected return on the equity.
- σ: the expected volatility of the commodity price.

When we deal with financial products, we need to take interest rates into account. Therefore, we introduce a discounted version of the equity price. Initially, let D_t denote the discounting factor determined by the ODE,

$$\mathrm{d}D_t = -r_t D_t \mathrm{d}t \qquad D_0 = 1. \tag{B.1.2}$$

Here r_t is the risk-free rate known to all market participants at time t, and hence D_t is deterministic with $D_t = e^{\int_0^t r(s)\mathrm{d}s}$ and a constant r, this simplifies to $D_t = e^{tr}$. We can now introduce the discounted stock price process governed by the Itô SDE:

$$\begin{aligned} \mathrm{d}\left(D_t S_t\right) &= D_t \mathrm{d}S_t + S_t \mathrm{d}D_t + \mathrm{d}S_t \mathrm{d}D_t \\ &= (\mu - r) D_t S_t \mathrm{d}t + \sigma D_t S_t \, \mathrm{d}B_t \\ &= \sigma D_t S_t \left(\frac{\mu - r}{\sigma} \mathrm{d}t + \, \mathrm{d}B_t \right). \end{aligned} \tag{B.1.3}$$

We now consider a transformation of the driving Brownian process of S_t such that:

$$\mathrm{d}B_t = \mathrm{d}\bar{B}_t - \theta_t \mathrm{d}t \tag{B.1.4}$$

As an example, one could consider a constant parameter $\theta_t = 3$, which would be a scaling in time that we would be able to simulate easily, see fig. B.2.

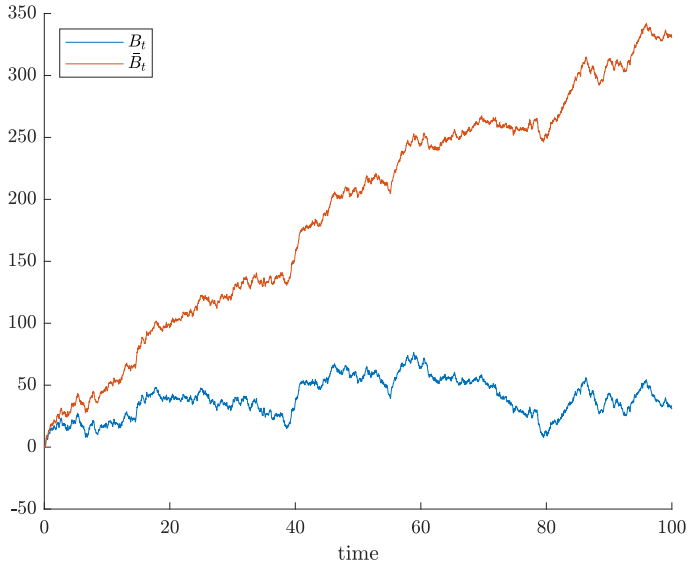


Figure B.2 – Simulated Paths of the B_t and $\bar{B}_t$ with $\theta_t = 3$

Notice in fig. B.2 that for every unique state of B_t there is a unique state of $\bar{B}_t$. Now in the general case, θ_t could be any function $\theta(t) : \mathbb{R}_+ \times \Omega \mapsto \mathbb{R}^s$ where s the dimension of the Brownian motion. In our example, with the geometric Brownian motion, we introduce $\mathrm{d}B_t = \mathrm{d}\bar{B}_t - \theta_t \mathrm{d}t$ and substitute this into eq. (B.1.1):

$$\mathrm{d}\left(D_t S_t\right) = \sigma D_t S_t \left[\frac{\mu - r}{\sigma}\mathrm{d}t + \left(\mathrm{d}\bar{B}_t - \theta_t \mathrm{d}t\right)\right] \tag{B.1.5}$$

The idea is now to choose θ_t in a clever way to remove the drift of the stochastic process and in that way make it a martingale process. Consider, therefore, the specific choice:

$$\theta_t = \frac{\mu - r}{\sigma} \tag{B.1.6}$$

Then in terms of the Brownian motion $\bar{B}_t$, the discounted price can be written as:

$$\mathrm{d}\left(D_t S_t\right) = \sigma D_t S_t \, \mathrm{d}\bar{B} \tag{B.1.7}$$

We see that using the transformation, we have removed the drift term from $\mathrm{d}\left(D_t S_t\right)$ which means it is now a martingale under this introduced transformation. This transformation is called the Girsanov transformation which we will define formally below. Additionally, the chosen $\theta_t = \frac{\mu - r}{\sigma}$ has the neat financial interpretation as the market price of risk for the considered stock.
It remains to understand in which space $\bar{B}_t$ is Brownian and how it relates to B_t as the postulate transformation is not trivial but is a direct application of the Girsanov theorem

Theorem: Girsanov II

Let $B(t, \omega)$ be a standard S-dimensional Brownian motion under the probability measure $\mathbb{P}$. Let $\boldsymbol{\theta} : \mathbb{R}_+ \times \Omega \mapsto \mathbb{R}^S$ be an adapted process such that the process Z defined as

$$Z(t, \omega) = \exp\left\{-\int_0^t \boldsymbol{\theta}^\top(t, \omega)\mathrm{d}\boldsymbol{B}(t, \omega) - \frac{1}{2}\int_0^t \boldsymbol{\theta}^\top(t, \omega)\boldsymbol{\theta}(t, \omega)\mathrm{d}t\right\}$$

satisfies $\mathrm{E}[Z(t,\omega)] = 1$. *Then the process*

$$\tilde{\boldsymbol{B}}(t,\omega) = \boldsymbol{B}(t,\omega) + \int_0^t \boldsymbol{\theta}(\omega,\tau)\mathrm{d}\tau$$

is a standard Brownian motion under the probability measure $\tilde{\mathbb{P}}$ *defined via the relation*

$$\mathrm{E}\left[\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}(\omega) \mid \mathcal{F}_t\right] = Z(t,\omega),$$

where F_t *is the natural filtration of the Brownian motion* $B(\omega,t)$.
So in the above, we implicitly used the Girsanov theorem and when we introduced the process $B(t,\omega)$ and the process θ_t to transform S_t, we actually introduced a new probability measure $\mathrm{d}\tilde{\mathbb{P}}$. The relation $\mathrm{E}\left[\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}(\omega) \mid \mathcal{F}_t\right] = Z(t,\omega)$ is often seen just as $\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}(\omega) = Z(t,\omega)$ in which case it is called the Radon-Nikodym derivative. In the following, we will also use $\mathbb{Q}$ instead of $\mathrm{d}\tilde{\mathbb{P}}$ for the risk-neutral pricing measure.

B.1.2 Risk-neutral Pricing

The risk-neutral measure is often used for the pricing of derivatives as we can utilize the Markov property under this measure. The most simple of which would be the pricing of a stock and how it related to the value of a future. Under the assumption of a complete market and assumption of no arbitrage possibilities, the fair price would be the discounted value of a future that can be bought at time t such that:

$$\begin{aligned} S_t &= \mathbb{E}^{\mathbb{Q}}\left[D_T S_T \middle| \mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{\int_0^T r(s)\mathrm{d}s} S_T \middle| \mathcal{F}_t\right] \\ &= e^{\int_t^T r(s)\mathrm{d}s}\mathbb{E}^{Q}\left[S_T \middle| \mathcal{F}_t\right] \end{aligned} \tag{B.1.8}$$

This is the classical relationship in mathematical finance where the spot price of a traded asset is the discounted future value under the risk-neutral measure. In the following, we will define the relationship in the context of the electricity market which differs slightly.

The Spot-Forward Relationship in Electricity Markets

Note that electricity is a non-traded asset as we cannot buy electricity and hold it as we can with equity asset classes such as stocks. If S_t were the price of a stock, then it would be costume to assume that the discounted future value, S_T, is martingale under $\mathbb{Q}$. This is perfectly reasonable as we would expect the stock to fall today t if the underlying company announces that it would have reduced production at time $t \leq T$. We assume that the market works efficiently and as stocks are tradeable, the information is priced into S_t. With that information priced in, we have no reason to believe that the discounted value should go in any specific direction i.e. S_t is a martingale under $\mathcal{F}_t$ and $\mathbb{Q}$. However, for electricity, the spot prices today, S_t, should not be affected by a planned shutdown of a power plant at time T, $t \geq T$. However, this should of cause be priced into the futures contract in the period covering T and hence that would increase. Essentially, this means that discounted future value need not be a martingale under $\mathbb{Q}$. Therefore, all measures $\mathbb{Q}$ which are equivalent to the real-world measure $\mathbb{P}$ are valid candidates [4] [3]. We will

find the measure that minimizes the distance between the observed forward prices and the calculated conditional expectation.
Consider, therefore, an electricity forward position with delivery at some time T and let S_T denote the anticipated price at time T. The payoff of such a position would be $S_T - F(t,T)$, and from classical financial theory, we know that the value of a derivative should be the present value of the future payoff under $\mathbb{Q}$ [3]:

$$e^{\int_t^T r(s)\mathrm{d}s}\mathbb{E}^Q\left[S_T - F(t,T)\middle|\mathcal{F}_t\right] = 0 \tag{B.1.9}$$

As $F(t,T)$ is the forward price set at price t, it must be based on all the current available information i.e. it must be adapted to $\mathbb{F}_t$ hence:

$$F\left(t,T\right) = \mathbb{E}^{\mathbb{Q}}\left[S_T\middle|\mathcal{F}_t\right]. \tag{B.1.10}$$

Note that we use $\mathbb{E}^{\mathbb{Q}}\left[\cdot\middle|\mathcal{F}_t\right]$ is the expected value under the historical filtration $\mathcal{F}_t = \sigma\left(S_u : 0 \leq u \leq t\right)$ as in def. 2.1, [4]. Additionally, we should note this is holds only under the equivalent measure, $\mathbb{Q}$. However, expectations under $\mathbb{Q}$ can be far from the expected prices in the real world. Therefore, as for most other commodity markets, the participants should be compensated for taking on the risk associated with holding the contract, *a risk premium*.

B.1.3 Risk Premium

The risk premium observed in the market is defined as the difference between the forward price and the expected spot price as introduced in (def. 4.2.2, [4]):

$$R\left(t,T\right) = F(t,T) - \mathbb{E}^{\mathbb{P}}\left[S_T\middle|\mathcal{F}_t\right]. \tag{B.1.11}$$

Here $F(t,T)$ is the forward price, and $\mathbb{E}^{\mathbb{P}}\left[S_T \mid \mathcal{F}_t\right]$ is the expected value of the spot price under the real-world probability and $\mathcal{F}_t$. $\mathbb{E}^{\mathbb{P}}\left[S_T\middle|\mathcal{F}_t\right]$ means that we will fit the model on historical spot prices and find the expected future spot price value without considering any need for risk compensation for the holders of the contract. If we calculated the risk premium empirically directly from the market, the equation would be:

$$\hat{R}\left(t,T\right) = \hat{F}(t,T) - \mathbb{E}^{\mathbb{P}}\left[S_T\middle|\mathcal{F}_t\right]. \tag{B.1.12}$$

We can also invoke the classical spot-forward relationship, eq. (B.1.10), and write the risk premium in terms of the expected spot price using the risk-neutral pricing measure,

$$R\left(t,T\right) = \mathbb{E}^{\mathbb{Q}}\left[S_T\middle|\mathcal{F}_t\right] - \mathbb{E}^{\mathbb{P}}\left[S_T\middle|\mathcal{F}_t\right]. \tag{B.1.13}$$

Recall that the measure $\mathbb{Q}$ need only be an equivalent measure to $\mathbb{P}$ for these electricity contracts. The difference between the two could e.g. be the simple drift we found as the market price of risk in eq. (B.1.6). For electricity futures, the sign of this risk premium could change with the time to delivery of the underlying. In [5], they analyze the sign of the risk premium and show how it changes. They hypothesize that the phenomena is explained by the market participants' appetite and need for risk diversification. For buyers

of electricity, the futures can be used to reduce the exposure to undesirable large price spikes and volatility. For the generators, they can hedge the cash flow of their generation assets as they lock in a price for future production. However, the time horizon for these hedging needs, is very different. The buyers want to hedge prices within the first two years while the generators would like to hedge production as far out in the future as possible. This gives means that electricity forward cures would often be in contango for 'shorter' maturities and be in backwardation for 'longer' dates maturities, see e.g. fig. A.1.

B.1.4 Filtrations

We have stipulated that $\mathcal{F}_t$ denotes a historical filtration of spot prices with no information about future important events. However, there could be filtration that could represent the information that market players have about future production and shutdowns. Therefore, introduce $\mathcal{H}_t$,

$$\mathcal{H}_t = \mathcal{F}_t \cup \sigma\left(S_{T_\Gamma}\right), \tag{B.1.14}$$

where $\sigma\left(S_{T_\Gamma}\right)$ contains relevant information about the underlying at time $t < T_\Gamma$ to perfectly describe what the price could be. Additionally, we introduce a more coarse filtration, $\mathcal{G}_t$, that could include some information about the possible level, K, of the underlying, e.g. $\mathcal{G}_t \subseteq \mathcal{F}_t \cup \sigma\left(\mathbb{1}_{S_{T_\Gamma} \geq K}\right)$. Indeed, we must have $\mathcal{F}_t \subseteq \mathcal{G}_t \subseteq \mathcal{H}_t$. We could formalize all of this through the theory of enlargement of filtrations and calculate the additional value that some participants have over others. We could at least calculate the difference between having only $\mathcal{F}_t$ and $\mathcal{G}_t$ which is called the *information premium*. In appendix C, we have developed some results based on [2] [4], however, we have decided to leave it out of the main report to limit our scope. Instead, we will move on and discuss how we will determine the future value of the spot price using a spot price model.

B.2 Spot Price Modelling

The dynamics of spot prices is characterized by three important components

1. strong mean reversion component
2. spikes caused by e.g. outages or favorable winds
3. a strong seasonal component

Based on these characteristics, we introduce the following model for the spot price, S_t

$$S_t = \Lambda\left(t\right) + X_t + Y_t, \tag{B.2.1}$$

where $\Lambda(t)$ is a deterministic function to capture the seasonal component in the spot price, X_t is an Ornstein-Uhlenbeck process defined by the SDE

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \sigma \mathrm{d}B_t, \tag{B.2.2}$$

with the mean reversion parameter $\alpha \in \mathbb{R}^+$, $\sigma \in \mathbb{R}^+$ is the volatility, and B_t is a standard Brownian motion. X_t will constitute the long time dynamics of the spot prices while Y_t is to represent the short spikes with

$$\mathrm{d}Y_t = -\beta Y_t \mathrm{d}t + dL_t, \tag{B.2.3}$$

with the mean reversion parameter $\beta \in \mathbb{R}^+$ and L_t is a square integrable Lévy process. Notice that X_t is independent of Y_t which makes analysis simpler and makes us able to study the seasonal, long term and spiky dynamics separately. In the following, we will introduce the Lévy process of interest and some useful analytical results that make us able to calculate the risk premium under this assumed model.

B.2.1 The Lévy Process of Interest

In our case, we will focus on a compound Poisson process with double-exponentially distributed jumps of L_t which in the litterateur is called the Kou model [6]. We define,

$$L_t = \sum_{i=1}^{N_t} D_i, \tag{B.2.4}$$

with $N_t \sim \text{Pois}(\lambda)$ e.g. N_t follows a Poisson process with intensity λ. As is commonly known, this means that the interarrival times of spikes are exponentially distributed e.g. if we denote the time of spikes as $\{t_1, t_2, \ldots, t_N\}$, then $t_i - t_{i-1} \sim \text{Exp}(\lambda)$ for $i = 2, 3, \ldots, N$. The jump sizes D_i will be i.i.d. and follow a double exponential distribution with density,

$$f_D(x) = p\eta_1 e^{-\eta_1 x}\mathbb{1}_{x\geq 0} + q\eta_2 e^{-\eta_2|x|}\mathbb{1}_{x\leq 0}, \tag{B.2.5}$$

with $p + q = 1$ and $\eta_1, \eta_2 \geq 0$. This model is useful as it has some useful analytical properties and results. Additionally, it allows us to model how negative and positive price spikes might be of different magnitudes. This is reasonable to use in the energy markets as the cause of negative and positive prices are very different; one is due to a surplus of energy and the because there is a great need for energy.

B.2.2 Integral form of the Spot Model

We can write the expression for the spot model in B.2.1 more explicitly by using that X_t is independent of Y_t hence we solve **??** and B.2.3 separately for some $u < t$ to obtain the integral form:

$$S_t = \Lambda(t) + e^{-\alpha(t-u)}X_u + \sigma\int_u^t e^{\alpha(t-s)}\mathrm{d}B_s + e^{-\beta(t-u)}Y_u + \int_u^t e^{\beta(t-s)}\mathrm{d}L_s. \tag{B.2.6}$$

When we calculate the price of a futures contract, we want to find the expectation under the assumption of the model above.

B.2.3 Pricing of Futures under the Assumed Model

With the introduced model for the spot price in eq. (B.2.6), we can now consider the expected price of a future at time t with delivery at time T. I.e. we want to evaluate,

$$F(t,T) = \mathbb{E}\left[S_T \middle| \mathcal{F}_t\right]. \tag{B.2.7}$$

The expectation of the deterministic function and the initial conditions are trivial. The expectation of the Itô integral is zero due to the martingale property of the Itô integral, however, the expectation of the Lêvy integral is not trivial. Consider, therefore, only $\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}L(s) \mid \mathcal{F}_t\right]$ and use Fubini's theorem to switch integration and expectation,

$$\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}L(s)\middle|\mathcal{F}_t\right] = \int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}\mathbb{E}\left[L(\mathrm{d}s)\mid\mathcal{F}_t\right]. \tag{B.2.8}$$

To calculate $\mathbb{E}\left[L(\mathrm{d}s)\mid\mathcal{F}_t\right]$, we need the log-moment generating function. Introduce the log-moment generating function of the process $\{L_t\}$ which by definition can be written as,

$$\psi_{L_{\mathrm{d}s}}(u) = \log\mathbb{E}\left[e^{uL_{\mathrm{d}s}}\right]. \tag{B.2.9}$$

The log-moment generating function has the convenient property,

$$\left.\frac{\mathrm{d}\psi_{L_{\mathrm{d}s}}(u)}{\mathrm{d}u}\right|_{u=0} = \mathbb{E}\left[L_{\mathrm{d}s}\right]. \tag{B.2.10}$$

For compound Poisson processes, we have analytical expressions for the log-moment generating function. From sec. 1.3 [7], we have:

$$\mathbb{E}\left[e^{uL_t}\right] = \exp\left(t\lambda\int_{\mathbb{R}}\left(e^{uy}-1\right)f_D(\mathrm{d}y)\right). \tag{B.2.11}$$

Where λ is the intensity of the Poisson process and f_D is the probability density of the jumps. With the Kou-model introduces above, and the property $\psi_{L_{\mathrm{d}s}}(u) = \log\mathbb{E}\left[e^{uL_t}\right]$, we can now derive $\psi_{L_{\mathrm{d}s}}(u)$. We have derived this in appendix A.2 and the ultimate result is:

$$\psi_{L_{\mathrm{d}s}}(u) = \lambda\left(\frac{qu}{u+\eta_2} - \frac{pu}{u-\eta_1}\right)\mathrm{d}t \tag{B.2.12}$$

$$\psi'_{L_{\mathrm{d}s}}(u) = \left[\frac{-\lambda p}{u-\eta_1} + \frac{\lambda p u}{(u-\eta_1)^2} + \frac{\lambda q}{u-\eta_2} + \frac{-\lambda q u}{(u-\eta_2)^2}\right]\mathrm{d}t \tag{B.2.13}$$

hence we now have the result:

$$\mathbb{E}\left[L_{\mathrm{d}s}\right] = \left.\frac{\mathrm{d}\psi_{L_{\mathrm{d}s}}(u)}{\mathrm{d}u}\right|_{u=0} = \lambda\left[\frac{p}{\eta_1} - \frac{q}{\eta_2}\right]\mathrm{d}t \tag{B.2.14}$$

With this result, we can now evaluate the expectation:

$$\begin{aligned}\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}L(s)\mid\mathcal{F}_t\right] &= \int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}\mathbb{E}\left[L(\mathrm{d}s)\mid\mathcal{F}_t\right] \\ &= \lambda\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)\int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}s \\ &= \frac{\lambda}{\beta}\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)\left(1-e^{-\beta(T-t)}\right)\end{aligned} \tag{B.2.15}$$

With this result, we can now write the futures price with this model as:

$$\begin{aligned}F(t,T) &= \mathbb{E}^{\mathbb{Q}}\left[S_T\middle|\mathcal{F}_t\right] \\ &= \Lambda(T) + e^{-\alpha(T-t)}X_t + e^{-\beta(T-t)}Y_t \\ &\quad + \frac{\lambda}{\beta}\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)\left(1-e^{-\beta(T-t)}\right)\end{aligned} \tag{B.2.16}$$

We will use the result above multiple times, however, to make this apply for electricity futures, we need to introduce a delivery period.

B.2.4 Futures with Delivery Period and Change of Measure

We need to extend our pricing framework to encompass the way electricity is traded in the futures market. The electricity contracts are traded for a delivery period $[T_1, T_2]$, $T_1, T_2 \in \mathbb{Z}$, and are most often settled financially such that the payoff for the person holding the contract is,

$$\sum_{t_i=T_1}^{T_2-1} S_{t_i} - (T_2 - T_1)\, F\left(t, T_1, T_2\right), \tag{B.2.17}$$

where S_{t_i} are the spot price for a specific day i, and $F\left(t, T_1, T_2\right)$ is the price of the future contract traded on day t for the delivery period $[T_1, T_2]$. We have $T_2 - 1$ as S_{T_2-1} is the spot price covering the last day in the interval. As an example, for the June-22 contract, then T_1 would be the first of June and T_2 the last day of June and hence essentially the participants trade the anticipated price of the flat average for all spot prices within the month.
If no arbitrage exists, then we must have, a risk-neutral pricing measure $\mathbb{Q}$, and a discounting rate r_t, then we can write the risk-neutral valuation as,

$$0 = e^{\int_{T_1}^{T_2} r(s)\mathrm{d}s}\mathbb{E}^{\mathbb{Q}}\left[\sum_{t_i=T_1}^{T_2-1} S_{t_i} - (T_2 - T_1)\, F\left(t, T_1, T_2\right), \middle| \mathcal{F}_t\right]. \tag{B.2.18}$$

where we have defined the discounting rate as in B.1.8.
We will for now assume that we consider forward contracts only. This means that the payment takes place at the end of the contract [1] and hence let $r \in \mathbb{R}$ denote the interest rate over the entire period then we can write:

$$0 = e^{-r(T_2-t)}\mathbb{E}^{\mathbb{Q}}\left[\sum_{t_i=T_1}^{T_2-1} S_{t_i} - (T_2 - T_1)\, F\left(t, T_1, T_2\right) \middle| \mathcal{F}_t\right]. \tag{B.2.19}$$

Additionally, it is natural to assume that the futures prices under $\mathbb{Q}$ are adapted to $\mathcal{F}_t$ because the price is based on the information in $\mathcal{F}_t$. Therefore, we take out what is known,

$$F^{\mathbb{Q}}_{\mathcal{F}_t}\left(t, T_1, T_2\right) = \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{T_2 - T_1}\sum_{t_i=T_1}^{T_2-1} S_{t_i} \middle| \mathcal{F}_t\right]. \tag{B.2.20}$$

In B.2.6, we defined a continuous time model for the daily prices. However, in the expectation above, we have a sum of discrete variables; one price for each day. However, if we assume that S_{t_i} is the expected value for the period it covers, then we could use the expected value of our continuous spot price model such that,

$$F^{\mathbb{Q}}_{\mathcal{F}_t}\left(t, T_1, T_2\right) = \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{T_2 - T_1}\sum_{t_i=T_1}^{T_2-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} S\left(u\right)\mathrm{d}u \mid \mathcal{F}_{t_i}\right] \,\middle|\, \mathcal{F}_t\right]. \tag{B.2.21}$$

As $\mathcal{F}_t \in \mathcal{F}_{t_i}$, we invoke the tower property to obtain,

$$F^{\mathbb{Q}}_{\mathcal{F}_t}\left(t, T_1, T_2\right) = \mathbb{E}^{\mathbb{Q}}\left[\int_{T_1}^{T_2} \frac{1}{T_2 - T_1} S(u)\,\mathrm{d}u \middle| \mathcal{F}_t\right]. \tag{B.2.22}$$

[1] If we considered futures, then the interest rate mattered as futures have daily settlements. This means that each day there is a payment that would have to be discounted with the rate r_t.

Note that with many of the same arguments, we could also derive the price of a future under the pricing measure $\mathbb{P}$,

$$F^{\mathbb{P}}_{\mathcal{F}_t}(t,T_1,T_2) = \mathbb{E}^{\mathbb{P}}\left[\int_{T_1}^{T_2} \frac{1}{T_2-T_1} S(u)\,\mathrm{d}u \middle| \mathcal{F}_t\right]. \tag{B.2.23}$$

Recall that $\mathbb{E}^{\mathbb{P}}[\,\cdot \mid \mathcal{F}_t]$ is to be understood as the expected price of spots within the delivery period without any considerations on the need to be compensated for the risk taken. Therefore, it is easier to find $F^{\mathbb{P}}_{\mathcal{F}_t}(t,T_1,T_2)$ hence we will start with that.
In both cases we introduce the following auxiliary functions to write all the solutions more compactly as in [3]:

$$\begin{aligned}
\bar{\alpha}(t,T_1,T_2) &= \begin{cases} -\frac{1}{\alpha}\left(e^{-\alpha(T_2-t)} - e^{-\alpha(T_1-t)}\right) & t \le T_1 \\ -\frac{1}{\alpha}\left(e^{-\alpha(T_2-t)} - 1\right) & t \ge T_1 \end{cases} \\
\bar{\beta}(t,T_1,T_2) &= \begin{cases} -\frac{1}{\beta}\left(e^{-\beta(T_2-t)} - e^{-\beta(T_1-t)}\right) & t \le T_1 \\ -\frac{1}{\beta}\left(e^{-\beta(T_2-t)} - 1\right) & t \ge T_1 \end{cases}
\end{aligned} \tag{B.2.24}$$

and additionally, we can define the functions

$$\begin{aligned}
\hat{\alpha}(t,T_1,T_2) &= \begin{cases} \frac{1}{\alpha}\left(T_2 - T_1 + \frac{1}{\alpha}\left(e^{-\alpha(T_2-t)} - e^{-\alpha(T_1-t)}\right)\right) & t \le T_1 \\ \frac{1}{\alpha}\left(T_2 - t + \frac{1}{\alpha}\left(e^{-\alpha(T_2-t)} - 1\right)\right) & t \ge T_1 \end{cases} \\
\hat{\beta}(t,T_1,T_2) &= \begin{cases} \frac{1}{\beta}\left(T_2 - T_1 + \frac{1}{\beta}\left(e^{-\beta(T_2-t)} - e^{-\beta(T_1-t)}\right)\right) & t \le T_1 \\ \frac{1}{\beta}\left(T_2 - t + \frac{1}{\beta}\left(e^{-\beta(T_2-t)} - 1\right)\right) & t \ge T_1 \end{cases}
\end{aligned} \tag{B.2.25}$$

Futures with Delivery Period under $\mathbb{P}$

Let $0 \le t \le T_1 < T_2$, then the forward price at t with delivery in $[T_1,T_2]$ and under the real-world measure $\mathbb{P}$ is then given by

$$\begin{aligned}
F^{\mathbb{P}}_{\mathcal{F}_t}(t,T_1,T_2) = \frac{1}{T_2-T_1}\Bigg(&\int_{T_1}^{T_2} \Lambda(u)\mathrm{d}u + \bar{\alpha}(t,T_1,T_2)\,X_t \\
&+\bar{\beta}(t,T_1,T_2)\,Y_t + \psi'_{L_1}(0)\hat{\beta}(t,T_1,T_2)\Bigg)
\end{aligned} \tag{B.2.26}$$

We can derive this directly from the model of the spot price where we have:

$$\begin{aligned}
F^{\mathbb{P}}_{\mathcal{F}_t}(t,T_1,T_2) &= \mathbb{E}^{\mathbb{P}}\left[\int_{T_1}^{T_2} \frac{1}{T_2-T_1} S(u)\,\mathrm{d}u \middle| \mathcal{F}_t\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\frac{1}{T_2-T_1}\int_{T_1}^{T_2} \Lambda(u) + X(u) + Y(u)\,\mathrm{d}u \middle| \mathcal{F}_t\right]
\end{aligned} \tag{B.2.27}$$

We can now consider each of the processes separately. The deterministic $\Lambda(t)$ is straightforward. For the case of $X(u)$, we write out the integral and appeal to the Fubini Theorem to obtain,

$$\begin{aligned}\int_{T_1}^{T_2} X(u)\,\mathrm{d}u &= \int_{T_1}^{T_2}\left(e^{-\alpha(u-t)}X(t) + \sigma\int_t^u e^{\alpha(u-s)}\mathrm{d}B_s\right)\mathrm{d}u \\ &= X(t)\left[-\frac{1}{\alpha}\left(e^{-\alpha(u-T_2)} - e^{-\alpha(u-T_1)}\right)\right] + \sigma\int_{T_1}^{T_2}\int_t^u e^{\alpha(u-s)}\mathrm{d}B_s\,\mathrm{d}u \\ &= X(t)\bar{\alpha}\left(t,T_1,T_2\right) + \sigma\int_t^u\int_{T_1}^{T_2} e^{\alpha(u-s)}\mathrm{d}u\,\mathrm{d}B_s \\ &= X(t)\bar{\alpha}\left(t,T_1,T_2\right) + \sigma\int_t^{T_2}\bar{\alpha}\left(t,T_1,T_2\right)\mathrm{d}B_s.\end{aligned}$$

As $X(t)$ is $\mathcal{F}_t$-measurable and due to the martingale property of the Itô integral, we find:

$$\mathbb{E}^{\mathbb{P}}\left[\int_{T_1}^{T_2} X(u)\,\mathrm{d}u\middle|\,\mathcal{F}_t\right] = X(t)\bar{\alpha}\left(t,T_1,T_2\right)$$

Likewise, we consider the expected value of $Y(t)$. However, as the expected size of the jumps is non-zero, we need to use the log-moment generating function,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[\int_{T_1}^{T_2} Y(u)\,\mathrm{d}u\,\mathcal{F}_t\right] &= \mathbb{E}^{\mathbb{P}}\left[\int_{T_1}^{T_2}\left(+e^{-\beta(t-u)}Y_u + \int_u^t e^{\beta(t-s)}\mathrm{d}L_s\right)\mathrm{d}u\,\mathcal{F}_t\right] \\ &= Y(t)\bar{\beta}(t,T_1,T_2) + \int_{T_1}^{T_2}\int_t^u e^{-\beta(u-s)}\mathbb{E}^{\mathbb{P}}\left[L_{\mathrm{d}s}\right]\mathrm{d}u \\ &= Y(t)\bar{\beta}(t,T_1,T_2) + \int_{T_1}^{T_2}\int_t^u e^{-\beta(u-s)}\psi'_{L_{\mathrm{d}s}}(0)\mathrm{d}s\mathrm{d}u \\ &= Y(t)\bar{\beta}(t,T_1,T_2) + \frac{1}{\beta}\psi'_{L_{\mathrm{d}s}}(0)\int_{T_1}^{T_2}\left(1-e^{-\beta(u-t)}\right)\mathrm{d}u \\ &= Y(t)\bar{\beta}(t,T_1,T_2) + \psi'_{L_{\mathrm{d}s}}(0)\bar{\beta}\left(t,T_1,T_2\right)\end{aligned}$$

Using the two results above, we see how we can write the price of the future under $\mathbb{P}$ as in eq. (B.2.26). We now consider how we can calculate the price of the future under the risk-neutral measure.

B.2.5 Change of Pricing Measure

In the following, we will cover how to find risk-neutral probabilities by introducing a risk-neutral pricing measure.

The risk-neutral Measure using the Girsanov and Esscher transformation

We first consider a risk-neutral measure for the X_t process governed by the Ornstein-Uhlenbeck process :

$$\mathrm{d}X_t = -\alpha X_t\mathrm{d}t + \sigma\mathrm{d}B_t. \tag{B.2.28}$$

Then we consider some transformation:

$$\mathrm{d}B_t = \mathrm{d}\bar{B}_t - \theta_t\mathrm{d}t. \tag{B.2.29}$$

In this case, it is convenient to consider a scaled function of θ_t i.e. $\theta_\beta(t) = -\frac{\theta(t)}{\sigma}$ such that $\mathrm{d}B_t = \mathrm{d}\bar{B}_t + \frac{\theta_\beta(t)}{\sigma}\mathrm{d}t$ and we obtain,

$$\begin{aligned} \mathrm{d}X_t &= -\alpha X_t \mathrm{d}t + \sigma \left[\mathrm{d}\bar{B}_t + \frac{\theta_\beta(t)}{\sigma} \mathrm{d}t \right] \\ &= \left(\theta_\beta(t) - \alpha X_t \right) \mathrm{d}t + \sigma \, \mathrm{d}\bar{B}_t. \end{aligned} \tag{B.2.30}$$

Already at this point, we see that for this process to be Martingale, we should choose $\theta_\beta(t) = \alpha X_t$. However, we will not do this directly as we are only looking for an equivalent measure. For now, we will just note that using Itô's lemma with $Y_t = h\left(X_t, t\right)$ where $h : \mathbb{R} \to \mathbb{R}$ and $h = e^t x$ we could easily see that this has the solution,

$$X_t = e^{-\alpha(t-s)} X_s + \int_s^t e^{-\alpha(t-u)} \theta_B(u) \mathrm{d}u + \sigma \int_s^t e^{-\alpha(t-u)} \mathrm{d}\bar{B}_t. \tag{B.2.31}$$

There exists a generalization of the Girsanov transformation for the Lévy case which is the Esscher transform[2],

$$Z_L(t) = \left. \frac{d\mathbb{Q}_L}{d\mathbb{P}} \right|_{\mathcal{F}_t} = \exp\left(\int_0^t \theta_L(s) dL_s - \int_0^t \psi_{L_1}\left(\theta_L(s)\right) ds \right)$$

where ψ_{L_1} is still the log-moment generating function, $\psi_{L_1}(u) = \log \mathbb{E}\left[e^{uL_1}\right]$, and $\theta_L(s)$ is the market price of jump-risk.
Consider an aggregated risk-neutral measure composed both of the market price of volatility risk and spike risk,

$$\mathbb{Q} = \mathbb{Q}_B \times \mathbb{Q}_L. \tag{B.2.32}$$

Then construct an aggregated vector of the measure change parameters $\theta(t) = \left(\theta(t)_B, \theta(t)_L\right)^\top$. Using this notation, we can do as in [3] and in the proof of proposition 3.4.3 in [2] and write the forward price under $\mathbb{Q}$.

Futures with Delivery Period under $\mathbb{Q}_\theta$

The price of a future with delivery in a period $[T_1, T_2]$ for the model eq. (B.2.6) under the equivalent measure $\mathbb{Q}$ is

$$\begin{aligned} F_{\mathcal{F}_t}^{\mathbb{Q}_\theta}\left(t, T_1, T_2\right) = \frac{1}{T_2 - T_1} \Bigg(& \int_{T_1}^{T_2} \Lambda(u) \mathrm{d}u + \bar{\alpha}\left(t, T_1, T_2\right) X_t + \bar{\beta}\left(t, T_1, T_2\right) Y_t \\ & + \int_t^{T_2} \theta_B(s) \bar{\alpha}\left(s, T_1, T_2\right) \mathrm{d}s + \int_t^{T_2} \psi'_{L_1}\left(\theta_L(s)\right) \bar{\beta}\left(s, T_1, T_2\right) \mathrm{d}s \Bigg). \end{aligned}$$

Here we have used the auxiliary functions $\bar{\alpha}$, $\bar{\beta}$ introduced in eq. (B.2.25). As stipulated before, the price under the risk-neutral probability need not be Martingal because electricity is a non-traded asset. Therefore, we can consider a simpler measure in which the parameter functions are constants which would simplify our forward price to be

$$\begin{aligned} F_{\mathcal{F}_t}^{\mathbb{Q}_\theta}\left(t, T_1, T_2\right) = \frac{1}{T_2 - T_1} \Bigg(& \int_{T_1}^{T_2} \Lambda(u) \mathrm{d}u \ + \bar{\alpha}\left(t, T_1, T_2\right) X_t + \bar{\beta}\left(t, T_1, T_2\right) Y_t \\ & + \theta_B \hat{\alpha}\left(t, T_1, T_2\right) + \psi'_{L_1}\left(\theta_L\right) \hat{\beta}\left(t, T_1, T_2\right) \Bigg). \end{aligned} \tag{B.2.33}$$

We will use the simplification above and find the vector of parameters that minimizes the distance squared distance to the observed futures.

[2] Some authors [8] refer to the transformation as *Girsanov Theorem I and II for Itô-Lévy Processes.*

Calculation of the Equivalent Measure $\mathbb{Q}_\theta$

We will find the measure $\mathbb{Q}_\theta$ from the observed forward prices $\hat{F}(t, T_1, T_2)$. Let $F_{\mathcal{F}}^{\mathbb{Q}_\theta}(t, T_1, T_2)$ be the future prices where we have tuned the parameters for the spot model but we need to tune the vector of measure change parameters θ. Additionally, consider a times series of daily closing prices for a specific future such at days $\{t_0, t_1, \ldots, t_N\}$ we have the series $\left\{\hat{F}(t_0, T_1, T_2), \hat{F}(t_1, T_1, T_2), \ldots, \hat{F}(t_N, T_1, T_2)\right\}$, then we want to find θ^*

$$\theta^* = \arg\min_\theta \quad \sum_{t_i=t_0}^{t_N} \left(\hat{F}(t, T_1, T_2) - F_{\mathcal{F}}^{\mathbb{Q}_\theta}(t, T_1, T_2)\right)^2. \tag{B.2.34}$$

The formula above is of course to be applied after we have found the parameters and functions that constitute the spot price model. In the following, we will cover how we will construct the spot price model.

B.2.6 Spot Price Model Building Approach

Inspired by the approach in [2], we will use a sequential model-building approach. If we let S_t denote the time series of daily spot prices in our dataset $\mathcal{D} = \{(t_1, s_1), (t_2, s_2), \ldots, (t_N, s_N)\}$. Then we will:

1. Determine the deterministic function, $\Lambda(t)$, and infer its parameters using S_t.
2. Filter the spikes from S_t using the normalized residuals from the model in $\Lambda(t)$. The spikes will be used to estimate the spike parameters Y_t.
3. Fine-tune the model in $\Lambda(t)$ without the spikes.
4. Find the parameters for $X(t)$ on the rest of the signal.

There exist methods that simultaneously estimate all the parameters, however, we use this approach to enable us to compare results with [2] and visually determine if the process seems to describe the spikes and variance in the data that we anticipated in the construction of the model. In the following, we will introduce the theory needed to apply the methods with an extended section of model parameter estimation for the SDE part of the model.

B.2.7 The Deterministic Function $\Lambda(t)$

The deterministic part of the price signal is to be understood as simple functions that capture dynamics in the spot market that are so characteristic that it can be considered the information available to all participants. We will consider a model with,

- A long-term general trend
- A weekly component with an intercept for each weekday as the prices are often different over the week.

In [2], they use sums of sin and cos with different periods to capture seasonality on multiple scales. However, they do not make any statistical arguments for the seasonalities. For instance, one would often see a yearly component as the prices usually follow the heating and cooling seasons of the year and therefore they added a term to capture that in [2]. However, it has been shown in e.g. [9] that there is no statistical evidence for a yearly seasonality in the German market in the period that we and [2] consider. Therefore, we

will start with a very simple model and only try to add the terms that we can see are needed in the residuals and check if they have statistical significance. We will start only with an intercept, trend, and weekly component,

$$\Lambda(t_i) = \beta t_i + \alpha_1 \mathbb{1}_{\{mon\}} + \alpha_2 \mathbb{1}_{\{tues\}} + \cdots + \alpha_7 \mathbb{1}_{\{sun\}} + \epsilon_{t_i}, \tag{B.2.35}$$

where $\epsilon_{t_i} \sim \mathcal{N}\left(0, \sigma^2\right)$. We will expand it slightly with a more elaborate model and change the covariance structure. We will also try to use a more generic framework called *Multiple Seasonal-Trend decomposition using Loes* to try to capture more of the variance and seasonal variance, see appendix E.1.
With the deterministic function, we can now move on and describe how we will find and determine the parameters for the diffusion part of our model **??**. The model is an Ornstein-Uhlenbeck process hence we will describe how numerical integration of SDEs and subsequently find the parameters.

C | Numerical Integration of SDEs

We can almost never solve stochastic differential equations analytically and hence must resort to approximate methods to obtain knowledge about the stochastic dynamical system. One way to obtain knowledge about the system is to simulate realizations of the possible paths a system makes given some initial condition. These methods are similar to the ones we know from numerical integration of deterministic systems where one sub-class of methods is called Runge-Kutta methods. We will here focus on a stochastic counterpart.

C.1 Euler-Maruyama

The following derivation is based on chapter 8 in [10]. A generic stochastic differential equation can be given as,

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + \mathbf{g}(\mathbf{x}, t)\mathrm{dB}(t), \tag{C.1.1}$$

where $x(t) \in \mathbb{R}^D$ and $B(t) \in \mathbb{R}^M$. We can also write eq. (C.1.1) on the integral form as,

$$\mathbf{x}(t) = \mathbf{x}\left(t_0\right) + \int_{t_0}^{t} \mathbf{f}(\mathbf{x}(\tau), \tau)\mathrm{d}\tau + \int_{t_0}^{t} \mathbf{g}(\mathbf{x}(\tau), \tau)\mathrm{dB}(\tau). \tag{C.1.2}$$

We can now apply Itô's formula to the drift, f, and the diffusion, g, and obtain,

$$\begin{aligned}
\mathrm{d}\mathbf{f}(\mathbf{x}(t), t) = &\frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial t}\ \mathrm{d}t + \sum_i \frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial x_i} f_i(\mathbf{x}(t), t)\mathrm{d}t \\
&+ \sum_i \frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial x_i} [\mathbf{g}(\mathbf{x}(t), t)\mathrm{dB}(\tau)]_i \\
&+ \frac{1}{2} \sum_{i,j} \frac{\partial^2 \mathbf{f}(\mathbf{x}(t), t)}{\partial x_i \partial x_j} \left[\mathbf{g}(\mathbf{x}(t), t)\mathbf{g}^\top(\mathbf{x}(t), t)\right]_{ij}\ \mathrm{d}t,
\end{aligned} \tag{C.1.3}$$

$$\begin{aligned}
\mathrm{d}\mathbf{g}(\mathbf{x}(t), t) = &\frac{\partial \mathbf{g}(\mathbf{x}(t), t)}{\partial t}\ \mathrm{d}t + \sum_i \frac{\partial \mathbf{g}(\mathbf{x}(t), t)}{\partial x_i} f_i(\mathbf{x}(t), t)\mathrm{d}t \\
&+ \sum_i \frac{\partial \mathbf{g}(\mathbf{x}(t), t)}{\partial x_i} [\mathbf{g}(\mathbf{x}(t), t)\mathrm{dB}(\tau)]_i \\
&+ \frac{1}{2} \sum_{i,j} \frac{\partial^2 \mathbf{g}(\mathbf{x}(t), t)}{\partial x_i \partial x_j} \left[\mathbf{g}(\mathbf{x}(t), t)\mathbf{g}^\top(\mathbf{x}(t), t)\right]_{ij}\ \mathrm{d}t.
\end{aligned} \tag{C.1.4}$$

We can now write eq. (C.1.3) and eq. (C.1.4) on integral form.

$$\begin{aligned}\mathbf{f}(\mathbf{x}(t),t) = \mathbf{f}\left(\mathbf{x}\left(t_0\right),t_0\right) &+ \int_{t_0}^{t} \frac{\partial \mathbf{f}(\mathbf{x}(\tau),\tau)}{\partial t}\ \mathrm{d}\tau \\ &+ \int_{t_0}^{t} \sum_i \frac{\partial \mathbf{f}(\mathbf{x}(\tau),\tau)}{\partial x_i} f_i(\mathbf{x}(\tau),\tau)\mathrm{d}\tau \\ &+ \int_{t_0}^{t} \sum_i \frac{\partial \mathbf{f}(\mathbf{x}(\tau),\tau)}{\partial x_i} [\mathbf{g}(\mathbf{x}(\tau),\tau)\mathrm{dB}(\tau)]_i \\ &+ \int_{t_0}^{t} \frac{1}{2}\sum_{i,j} \frac{\partial^2 \mathbf{f}(\mathbf{x}(\tau),\tau)}{\partial x_i \partial x_j} \left[\mathbf{g}(\mathbf{x}(\tau),\tau)\mathbf{g}^\top(\mathbf{x}(\tau),\tau)\right]_{ij}\ \mathrm{d}\tau\end{aligned} \tag{C.1.5}$$

$$\begin{aligned}\mathbf{g}(\mathbf{x}(t),t) = \mathbf{g}\left(\mathbf{x}\left(t_0\right),t_0\right) &+ \int_{t_0}^{t} \frac{\partial \mathbf{g}(\mathbf{x}(\tau),\tau)}{\partial t}\ \mathrm{d}\tau \\ &+ \int_{t_0}^{t} \sum_i \frac{\partial \mathbf{g}(\mathbf{x}(\tau),\tau)}{\partial x_i} f_i(\mathbf{x}(\tau),\tau)\mathrm{d}\tau \\ &+ \int_{t_0}^{t} \sum_i \frac{\partial \mathbf{g}(\mathbf{x}(\tau),\tau)}{\partial x_i} [\mathbf{g}(\mathbf{x}(\tau),\tau)\mathrm{dB}(\tau)]_i \\ &+ \int_{t_0}^{t} \frac{1}{2}\sum_{i,j} \frac{\partial^2 \mathbf{g}(\mathbf{x}(\tau),\tau)}{\partial x_i \partial x_j} \left[\mathbf{g}(\mathbf{x}(\tau),\tau)\mathbf{g}^\top(\mathbf{x}(\tau),\tau)\right]_{ij}\ \mathrm{d}\tau\end{aligned} \tag{C.1.6}$$

We now define the following two operators to simplify the notation.

$$\begin{aligned}L_t(\cdot) &= \frac{\partial(\cdot)}{\partial t} + \sum_i \frac{\partial(\cdot)}{\partial x_i} f_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2(\cdot)}{\partial x_i \partial x_j}\left[\mathbf{g}\mathbf{g}^\top\right]_{ij}, \\ L_{\mathrm{B},j}(\cdot) &= \sum_i \frac{\partial(\cdot)}{\partial x_i} g_{ij}, \quad \text{for } j = 1,2,\ldots,M,\end{aligned} \tag{C.1.7}$$

The new operators reduces eq. (C.1.5) and eq. (C.1.6) to

$$\begin{aligned}\mathbf{f}(\mathbf{x}(t),t) =& \mathbf{f}\left(\mathbf{x}\left(t_0\right),t_0\right) + \int_{t_0}^{t} L_t \mathbf{f}(\mathbf{x}(\tau),\tau)\mathrm{d}\tau \\ &+ \sum_j \int_{t_0}^{t} L_{\mathrm{B},j}\mathbf{f}(\mathbf{x}(\tau),\tau)\mathrm{dB}_j(\tau),\end{aligned} \tag{C.1.8}$$

$$\begin{aligned}\mathbf{g}(\mathbf{x}(t),t) =& \mathbf{g}\left(\mathbf{x}\left(t_0\right),t_0\right) + \int_{t_0}^{t} L_t \mathbf{g}(\mathbf{x}(\tau),\tau)\mathrm{d}\tau \\ &+ \sum_j \int_{t_0}^{t} L_{\mathrm{B},j}\mathbf{g}(\mathbf{x}(\tau),\tau)\mathrm{dB}_j(\tau),\end{aligned} \tag{C.1.9}$$

We now substitute back into eq. (C.1.2).

$$\begin{aligned}\mathbf{x}(t) = \mathbf{x}\left(t_0\right) + \mathbf{f}\left(\mathbf{x}\left(t_0\right), t_0\right)\left(t - t_0\right) + \mathbf{g}\left(\mathbf{x}\left(t_0\right), t_0\right)\left(\mathrm{B}(t) - \mathrm{B}\left(t_0\right)\right) \\ + \int_{t_0}^{t}\int_{t_0}^{\tau} L_t\mathbf{f}(\mathbf{x}(\tau),\tau)\mathrm{d}\tau\mathrm{d}\tau + \sum_j \int_{t_0}^{t}\int_{t_0}^{\tau} L_{\mathrm{B},j}\mathbf{f}(\mathbf{x}(\tau),\tau)\mathrm{dB}_j(\tau)\mathrm{d}\tau \\ + \int_{t_0}^{t}\int_{t_0}^{\tau} L_t\mathbf{g}(\mathbf{x}(\tau),\tau)\mathrm{d}\tau\mathrm{dB}(\tau) \\ + \sum_j \int_{t_0}^{t}\int_{t_0}^{\tau} L_{\mathrm{B},j}\mathbf{g}(\mathbf{x}(\tau),\tau)\mathrm{dB}_j(\tau)\mathrm{dB}(\tau).\end{aligned} \tag{C.1.10}$$

We now assume that there is only a minor contribution from each of the integrals above and collect the contributions from all the involved terms in $\mathbf{r}(t)$

$$\mathbf{x}(t) = \mathbf{x}\left(t_0\right) + \mathbf{f}\left(\mathbf{x}\left(t_0\right), t_0\right)\left(t - t_0\right) + \mathbf{g}\left(\mathbf{x}\left(t_0\right), t_0\right)\left(\mathrm{B}(t) - \mathrm{B}\left(t_0\right)\right) + \mathbf{r}(t). \tag{C.1.11}$$

We recognize eq. (C.1.11) as the Euler-Maruyama scheme but before we talk more about this scheme, we need to understand convergence for stochastic schemes.

C.1.1 Convergence of Stochastic Schemes

We have two notions of convergence for stochastic differential equations. We will refer to a specific realization as x and the random approximation scheme as X. First, we have what is known as strong convergence. It is described in section 5 in [11] and given as

$$\mathbb{E}\left[|x - X|\right] \leq Ch^{\gamma}. \tag{C.1.12}$$

and weak convergence is given as

$$\left|\mathbb{E}\left[x\right] - \mathbb{E}\left[X\right]\right| \leq Ch^{\delta}. \tag{C.1.13}$$

In the context of numerical schemes for stochastic differential equations, strong and weak convergence are distinct concepts. Weak convergence suggests that the simulated average path is accurate to within a term of h^{δ}, whereas strong convergence guarantees that every individual path is precise to the degree of h^{γ}. The Euler-Maruyama method has been shown to have a strong order of convergence of $\gamma = 1/2$ and a weak order of convergence of $\delta = 1$. The relatively low strong order of convergence is attributed to the presence of the term $\mathrm{dB}_j(\tau)\mathrm{dB}(\tau)$ in the residual, which upon integration results in a term with $\mathrm{dB}(\tau)$, which is only of order $\mathrm{d}t^{1/2}$. However, by expanding this term, it is possible to increase the strong order to one.

C.1.2 Methods with Higher Order than Euler Maruyama

To increase the order, similar to what was done in eq. (C.1.9), we can perform the same type of expansion for the term $L_{\mathrm{B},j}\mathbf{g}(\mathbf{x}(\tau),\tau)$, as outlined in chapter 8 of [10].

$$\begin{aligned}L_{\mathrm{B},j}\mathbf{g}(\mathbf{x}(t),t) = L_{\mathrm{B},j}\mathbf{g}\left(\mathbf{x}\left(t_0\right),t_0\right) + \int_{t_0}^{t} L_t L_{\mathrm{B},j}\mathbf{g}(\mathbf{x}(t),t)\mathrm{d}t \\ + \sum_j \int_{t_0}^{t} L_{\mathrm{B},j}^2\mathbf{g}(\mathbf{x}(t),t)\mathrm{dB}_j(\tau)\end{aligned} \tag{C.1.14}$$

We can now substitute this into eq. (C.1.10) and obtain.

$$\begin{aligned}\mathbf{x}(t) =&\mathbf{x}\left(t_0\right)+\mathbf{f}\left(\mathbf{x}\left(t_0\right),t_0\right)\left(t-t_0\right)+\mathbf{g}\left(\mathbf{x}\left(t_0\right),t_0\right)\left(\mathrm{B}(t)-\mathrm{B}\left(t_0\right)\right)\\&+\sum_j L_{\mathrm{B},j}\mathbf{g}\left(\mathbf{x}\left(t_0\right),t_0\right)\int_{t_0}^{t}\int_{t_0}^{\tau}\mathrm{dB}_j(\tau)\mathrm{dB}(\tau)+\mathbf{r}(t)\end{aligned} \tag{C.1.15}$$

This is now the method called Milstein's method which has a strong order of 1. It though has two issues. The first is the iterated Itô integral,

$$\int_{t_0}^{t}\int_{t_0}^{\tau}\mathrm{dB}_j(\tau)\mathrm{dB}(\tau). \tag{C.1.16}$$

Such integrals pose a significant challenge and as a result, the Milstein method cannot be easily extended to SDEs with arbitrary noise structures. Some noise structures though allow the iterated Itô integral to be solved analytically. One such case is the scalar noise case which gives the following version of Milstein's method.

$$\begin{aligned}x\left(t_{k+1}\right)=x\left(t_k\right)&+f\left(x\left(t_k\right),t_k\right)\Delta t+g\left(x\left(t_k\right),t_k\right)\Delta\mathrm{B}_k\\&+\frac{1}{2}\frac{\partial g\left(x\left(t_k\right),t_k\right)}{\partial x}g\left(x\left(t_k\right),t_k\right)\left(\Delta\mathrm{B}_k^2-\Delta t\right).\end{aligned} \tag{C.1.17}$$

Here, we encounter the second issue. The Milstein method belongs to a group of techniques referred to as Itô-Taylor Series-based methods, which necessitate explicit derivatives of the drift and diffusion functions. This problem is less complex to address than the iterated Itô integrals and in the next sectionm we will see a class of methods that do not need explicit derivatives.

C.2 Stochastic Runge-Kutta Methods

The family of Runge-Kutta methods gives us a class of methods that avoids explicit derivatives of the drift and diffusion functions. For the deterministic case, an s-stage Runge-Kutta method is given by the following form,

$$\begin{aligned}\mathbf{H}_1&=X_n+\mathbf{f}\left(t_n+c_1\Delta t,X_n+\Delta t\sum_{\ell=1}^{s}a_{1\ell}\mathbf{H}_\ell\right)\\&\vdots\\\mathbf{H}_s&=X_n+\mathbf{f}\left(t_n+c_s\Delta t,X_n+\Delta t\sum_{\ell=1}^{s}a_{s\ell}\mathbf{H}_\ell\right)\end{aligned} \tag{C.2.1}$$

$$X_{n+1}=X_n+\Delta t\sum_{i=1}^{s}b_i\mathbf{H}_i, \tag{C.2.2}$$

where all coefficients a_{il}, b_i, $c_i\in\mathbb{R}$. The coefficients can then be arranged in a Butcher tableau which simplifies the development and analysis of these methods greatly. Such a tableau is shown below.

$$\begin{array}{c|c}\mathbf{c}&\mathbf{A}\\\hline&\mathbf{b}^T\end{array}:=\begin{array}{c|ccc}c_1&a_{1,1}&\cdots&a_{1,s}\\\vdots&\vdots&\ddots&\vdots\\c_s&a_{s,1}&\cdots&a_{s,s}\\\hline&b_1&\cdots&b_s\end{array} \tag{C.2.3}$$

For the stochastic case, it is though not as simple due to the iterated Itô integrals. Hence how methods are stated differs depending on the noise structure.
We will in the following first analyze the case of additive noise and then expand the scope to a scalar noise structure.

C.2.1 SRK with Additive Noise and Strong Order 1.5

We can write up the structure for stochastic Runge Kutta methods in a similar way as for the deterministic Runge Kutta methods given in eq. (C.2.1) and eq. (C.2.2). For stochastic Runge Kutta methods with additive noise and strong order 1.5, we have the following structure.

$$\begin{aligned} X_{n+1} = X_n &+ \sum_{i=1}^{s} \alpha_i f\left(t_n + c_i^{(0)}\Delta t, H_i^{(0)}\right)\Delta t \\ &+ \sum_{i=1}^{s}\left(\beta_i^{(1)} I_{(k)} + \beta_i^{(2)}\frac{I_{(1,0)}}{\Delta t}\right) g\left(t_n + c_i^{(1)}\Delta t\right) \end{aligned} \tag{C.2.4}$$

for time steps $n = 0, 1, \ldots, N-1$ and with the intermediate stages,

$$H_i^{(0)} = Y_n + \sum_{j=1}^{s} A_{ij}^{(0)} f\left(t_n + c_j^{(0)}\Delta t, H_j^{(0)}\right)\Delta t + \sum_{j=1}^{s} B_{ij}^{(0)} g\left(t_n + c_j^{(1)}\Delta t\right)\frac{I_{(1,0)}}{\Delta t},$$

where $i = 0, 1, \ldots, s$. With this noise structure, the iterated Itô integrals are still analytically feasible and can be written the following way,

$$\begin{aligned} I_{(1)} &= \Delta \mathrm{B}(t) \sim \mathrm{N}(0, \Delta t) \\ I_{(1,0)} &= \frac{1}{2}\Delta t(I_{(1)} - 1/\sqrt{3}\xi), \quad \xi \sim \mathrm{N}(0, \Delta t)\,. \end{aligned} \tag{C.2.5}$$

In deterministic Runge-Kutta theory, we can use the Butcher tableau with rooted tree analysis, [12]. This we can use to set up order conditions from which we can derive coefficients for an integration scheme of a certain order. Due to the iterated Itô integrals this is not as straightforward in the stochastic case. We though still have a similar concept called colored rooted tree analysis, [13], which we can use to derive order conditions in the stochastic setting. For methods with additive noise, the Butcher tableau is of the following structure,

$$\begin{array}{c|c|c|c} c^{(0)} & A^{(0)} & B^{(0)} & c^{(1)} \\ \hline & \alpha^T & \beta^{(1)^T} & \beta^{(2)^T} \end{array}. \tag{C.2.6}$$

In [14] they derived order conditions for methods of strong order 1.5 with additive noise and found that for the method to be of order 1 the following conditions must be met.

$$\textbf{1. } \alpha^T e = 1 \quad \textbf{2. } \beta^{(1)^T} e = 1 \quad \textbf{3. } \beta^{(2)^T} e = 0 \tag{C.2.7}$$

and the additional conditions for order 1.5:

$$\begin{aligned} &\textbf{1. } \alpha^T B^{(0)} e = 1 \quad \textbf{2. } \alpha^T A^{(0)} e = \tfrac{1}{2} \quad \textbf{3. } \alpha^T \left(B^{(0)} e\right)^2 = \tfrac{3}{2} \\ &\textbf{4. } \beta^{(1)^T} c^{(1)} = 1 \quad \textbf{5. } \beta^{(2)^T} c^{(1)} = -1 \end{aligned} \tag{C.2.8}$$

From these conditions they, [14], proposed the following strong order 1.5 scheme known as SRA3:

$$\begin{array}{c|ccc|ccc|ccc} 0 & & & & & & & 1 & & \\ 1 & 1 & & & 0 & & & 0 & & \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & & 1 & \frac{1}{2} & & 0 & & \\ \hline & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & 1 & 0 & 0 & 1 & -1 & 0 \end{array} \tag{C.2.9}$$

C.2.2 SRK with Scalar Noise and Strong Order 1.5

We now expand the noise structure to also include scalar noise. The strong order we keep at 1.5. Such methods have the following structure.

$$\begin{aligned} X_{k+1} = X_k &+ \sum_{i=1}^{s} \alpha_i \mathbf{f}\left(t_k + c_i^{(0)}\Delta t, H_i^{(0)}\right)\Delta t \\ &+ \sum_{i=1}^{s}\left(\beta_i^{(1)} I_{(1)} + \beta_i^{(2)}\frac{I_{(1,1)}}{\sqrt{\Delta t}} + \beta_i^{(3)}\frac{I_{(1,0)}}{\Delta t} + \beta_i^{(4)}\frac{I_{(1,1,1)}}{\Delta t}\right)\mathbf{g}\left(t_k + c_i^{(1)}\Delta t, H_i^{(1)}\right) \end{aligned} \tag{C.2.10}$$

for time steps $n = 0, 1, \ldots, N-1$ with the intermediate stages:

$$\begin{aligned} H_i^{(0)} &= X_k + \sum_{j=1}^{s} A_{ij}^{(0)}\mathbf{f}\left(t_k + c_j^{(0)}\Delta t, H_j^{(0)}\right)\Delta t + \sum_{j=1}^{o} B_{ij}^{(0)}\mathbf{g}\left(t_k + c_j^{(1)}\Delta t, H_j^{(1)}\right)\frac{I_{(1,0)}}{\Delta t}, \\ H_i^{(1)} &= X_k + \sum_{j=1}^{s} A_{ij}^{(1)}\mathbf{f}\left(t_k + c_j^{(0)}\Delta t, H_j^{(0)}\right)\Delta t + \sum_{j=1}^{s} B_{ij}^{(1)}\mathbf{g}\left(t_k + c_j^{(1)}\Delta t, H_j^{(1)}\right)\sqrt{\Delta t} \end{aligned} \tag{C.2.11}$$

where $i = 0, 1, \ldots, s$. The iterated Itô integrals are still analytically tractable in this noise structure and are given as follows:

$$I_{(1)} = \mathrm{dB}(t) \tag{C.2.12}$$

$$I_{(1,1)} = \frac{1}{2}(I_{(1)}^2 - \Delta t) \tag{C.2.13}$$

$$I_{(1,0)} = \frac{1}{2}\Delta t I_{(1)} - \Delta t\frac{1}{2\sqrt{3}}\xi, \quad \text{where } \xi \sim \mathcal{N}(0, \Delta t) \tag{C.2.14}$$

$$I_{(1,1,1)} = \frac{1}{6}(I_{(1)}^3 - 3\Delta t I_{(1)}) \tag{C.2.15}$$

The Butcher tableau is expanded to the following form.

$$\begin{array}{c|c|cc} c^{(0)} & A^{(0)} & B^{(0)} & \\ \hline c^{(1)} & A^{(1)} & B^{(1)} & \\ \hline & \alpha^T & \beta^{(1)^T} & \beta^{(2)^T} \\ \hline & & \beta^{(3)^T} & \beta^{(4)^T} \end{array} \tag{C.2.16}$$

In [14] they also derived order conditions for methods with scalar noise. We can now define a set of conditions that must be satisfied for the model to be of a specific order. We start with the condition needed for the methods to be of order 0.5,

$$\begin{array}{lll} \textbf{1. } \alpha^T e = 1 & \textbf{2. } \beta^{(1)^T} e = 1 & \textbf{3. } \beta^{(2)^T} e = 0 \\ \textbf{4. } \beta^{(3)^T} e = 0 & & \textbf{5. } \beta^{(4)^T} e = 0, \end{array} \tag{C.2.17}$$

additionally, for order 1,

$$\begin{array}{ll} \textbf{1.}\ \beta^{(1)^T} B^{(1)} e = 0 & \textbf{2.}\ \beta^{(3)^T} B^{(1)} e = 0 \\ \textbf{3.}\ \beta^{(2)^T} B^{(1)} e = 1 & \textbf{4.}\ \beta^{(4)^T} B^{(1)} e = 0, \end{array} \tag{C.2.18}$$

and lastly for order 1.5,

1. $\alpha^T A^{(0)} e = \frac{1}{2}$	**2.** $\alpha^T B^{(0)} e = 1$	**3.** $\alpha^T \left(B^{(0)} e\right)^2 = \frac{3}{2}$
4. $\beta^{(1)^T} A^{(1)} e = 1$	**5.** $\beta^{(2)^T} A^{(1)} e = 0$	**6.** $\beta^{(3)^T} A^{(1)} e = -1$
7. $\beta^{(4)^T} A^{(1)} e = 0$	**8.** $\beta^{(1)^T} \left(B^{(1)} e\right)^2 = 1$	**9.** $\beta^{(2)^T} \left(B^{(1)} e\right)^2 = 0$
10. $\beta^{(3)^T} \left(B^{(1)} e\right)^2 = -1$	**11.** $\beta^{(4)^T} \left(B^{(1)} e\right)^2 = 2$	**12.** $\beta^{(1)^T} \left(B^{(1)} \left(B^{(1)} e\right)\right) = 0$
13. $\beta^{(2)^T} \left(B^{(1)} \left(B^{(1)} e\right)\right) = 0$	**14.** $\beta^{(3)^T} \left(B^{(1)} \left(B^{(1)} e\right)\right) = 0$	**15.** $\beta^{(4)^T} \left(B^{(1)} \left(B^{(1)} e\right)\right) = 1$

From these conditions they, [14], proposed the following strong order 1.5 scheme known as SRIW1 which has the following tableau,

$$\begin{array}{c|cccc|cccc|cccc}
0 & & & & & & & & & & & & \\
1 & 1 & & & & 0 & & & & & & & \\
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & & & 1 & \frac{1}{2} & & & & & & \\
0 & 0 & 0 & 0 & & 0 & 0 & 0 & & & & & \\
\hline
0 & & & & & & & & & & & & \\
\frac{1}{4} & \frac{1}{4} & & & & -\frac{1}{2} & & & & & & & \\
1 & 1 & 0 & & & 1 & 0 & & & & & & \\
\frac{1}{4} & 0 & 0 & \frac{1}{4} & & 2 & -1 & \frac{1}{2} & & & & & \\
\hline
 & \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & 0 & -1 & \frac{4}{3} & \frac{2}{3} & 0 & 1 & -\frac{4}{3} & \frac{1}{3} & 0 \\
\hline
 & & & & & 2 & -\frac{4}{3} & -\frac{2}{3} & 0 & -2 & \frac{5}{3} & -\frac{2}{3} & 1
\end{array}$$

Figure C.1 – The SRIW1 method with strong order 1.5 and weak order 3.

C.2.3 Numerical Experiments

We will now test the SRIW1 method and compare two other methods. A similar method for scalar noise of only strong order 1 is given in appendix B.1, and the Euler-Maruyama method which has strong order 0.5. We will test on a 1-dimensional geometric Brownian motion which is given as

$$\mathrm{dX}_t = \mu \mathrm{X}_t \mathrm{dt} + \sigma \mathrm{X}_t \mathrm{dB}_t \tag{C.2.19}$$

The parameter μ can be interpreted as "percentage drift" and σ can be interpreted as "percentage volatility". The solution to eq. (C.2.19) is given by

$$\mathrm{X}_t = \mathrm{X}_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma \mathrm{B}_t\right) \tag{C.2.20}$$

To test the Euler-Maruyama, the SRK of strong order 1.0, and the SRK of strong order 1.5, we sample a Brownian motion very finely and then sub-sample the same path with a cruder step size as illustrated in fig. C.2.

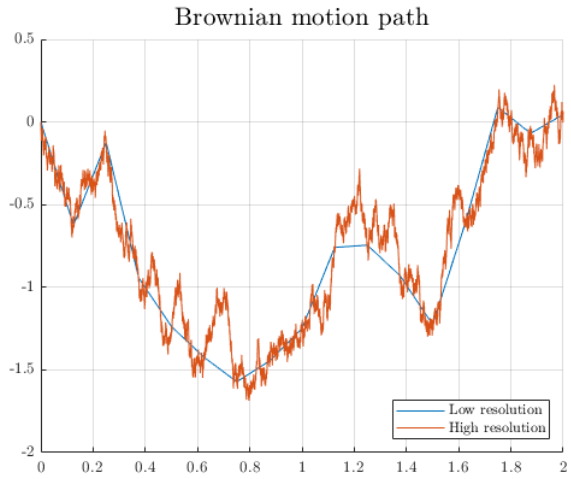


Figure C.2

We can utilize the high-resolution Brownian motion to plot the analytical solution and then use the low-resolution path to evaluate the numerical methods. As shown in fig. C.3, a step size of 2^{-3} was used for the numerical methods, while the analytical solution used 2^{-14}. We observe that the Euler-Maruyama method has a substantial error at the end, and the stronger the order, the smaller the error is.

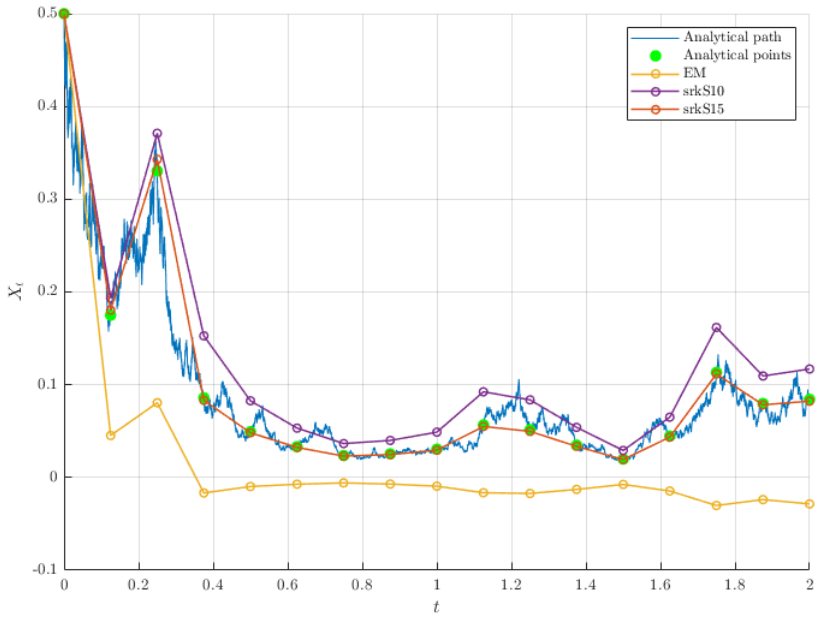


Figure C.3

We now simulate the stochastic Runge-Kutta methods for 1000 different realizations of Brownian motion using different step sizes. We then compare the endpoint, X_T, with the analytical solution and plot the log error of the different methods as a function of the step size. In fig. C.4, we see that the higher the order, the faster the error decreases with the step size. The slope of the Euler-Maruyama log error is 0.586 which is close to 0.5 as

expected. Similarly, the slope of the SRKs10 method is 0.958 and 1.477 for the SRKs15 method, also called SRIW1.

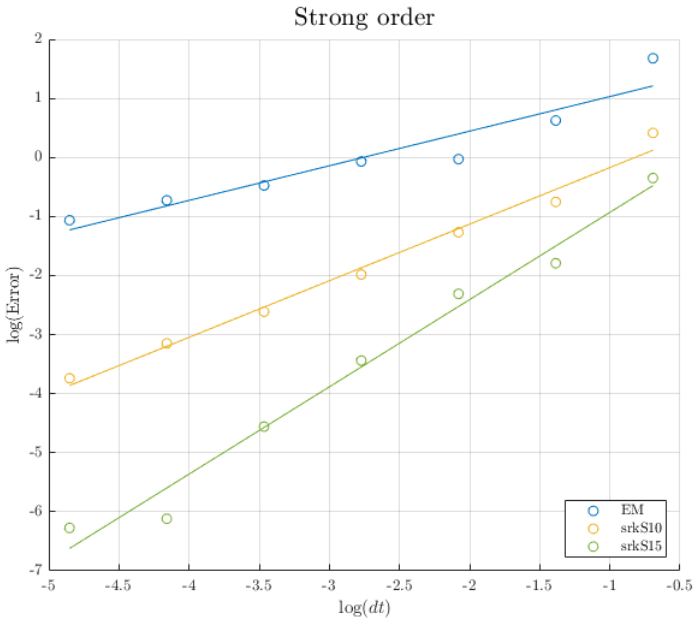


Figure C.4

C.2.4 Stability Analysis

Previously, we have only discussed convergence, however, when working with numerical schemes for differential equations, it is also important to consider whether the scheme inherits the stability properties of the original differential equation. To provide context for this analysis of our schemes for stochastic differential equations, we will first examine deterministic ODEs. According to [12], stability is defined as follows.

Assume an initial value problem (IVP) is given as

$$\dot{x}(t) = f(x(t)), \quad x(t_0) = x_0. \tag{C.2.21}$$

Then the stability of the system is defined as follows:

1. **asymptotically stable** if $\|\mathbf{x}(t)\| \to 0$ as $t \to \infty$ for all initial values $\mathbf{x}_0 \in \mathbb{R}^d$;
2. **stable** if there is a constant C (independent of t and $\mathbf{x}_0$) such that $\|\mathbf{x}(t)\| < C \|\mathbf{x}_0\|$ holds for all $t \geq t_0$ and $\mathbf{x}_0 \in \mathbb{R}^d$
3. **unstable**, otherwise.

It is though often very challenging to say anything about the stability of the discretization of such a general system as eq. (C.2.21). One hence limits the scope to a linear test equation as the one stated below where G is a matrix.

$$\dot{x}(t) = Gx(t), \quad x(t_0) = x_0, \tag{C.2.22}$$

The stability of eq. (C.2.22) is then as follows.

1. eq. (C.2.22) is *asymptotically stable* if and only if all eigenvalues λ of G satisfy $\mathrm{Re}(\lambda) < 0$.

2. The IVP given in eq. (C.2.22) is *stable* if and only if all eigenvalues λ of G satisfy $\mathrm{Re}(\lambda) \leq 0$ and if every eigenvalue λ with $\mathrm{Re}(\lambda) = 0$ has the same algebraic and geometric multiplicity.

The previous is proved in [12] under theorem 3.2. We can now use a given Runge Kutta method to discretize the solution of eq. (C.2.22). This we can write as a linear map of the linear IVP given in eq. (C.2.22).

$$x_{n+1} = R(hG)x_n \tag{C.2.23}$$

The stability of such a discrete map is then defined as follows:

1. **asymptotically stable** if $\|\mathbf{x}_k\| \to 0$ as $k \to \infty$ for all initial values $\mathbf{x}_0 \in \mathbb{R}^d$

2. **stable** if there is a constant C (independent of t and $\mathbf{x}_0$) such that $\|\mathbf{x}_k\| < C\,\|\mathbf{x}_0\|$ holds for all $k \geq 0$ and $\mathbf{x}_0 \in \mathbb{R}^d$

3. **unstable**, otherwise.

As for the linear continuous system, we can also say a lot about the stability of the linear map given in eq. (C.2.23).

1. The linear discrete map in eq. (C.2.23) is *asymptotically stable* if and only if all eigenvalues λ of $R(hG)$ satisfy $|\lambda| < 1$.

2. The linear discrete map in eq. (C.2.23) is *stable* if and only if all eigenvalues λ of $R(hG)$ satisfy $|\lambda| \leq 1$ and if every eigenvalue λ with $|\lambda| = 1$ has the same algebraic and geometric multiplicity.

The previous is proved in [12] under theorem 3.3.

Hence what we now want to answer is if our discrete map inherits the same stability as the continuous counterpart. To answer this, we need to investigate whether the eigenvalues of $R(hG)$ have absolute magnitude less than 1. We note that the eigenvalues of $R(hG)$ are given by $R(h\lambda)$ for every eigenvalue λ of G. We can hence define the stability region as

$$\mathcal{S} := \{h\lambda \in \mathbb{C} : |R(h\lambda)| \leq 1\}. \tag{C.2.24}$$

Path-wise Stability in the Stochastic Setting

To extend the concepts of stability from deterministic systems to the stochastic setting, we need to define stability in a probabilistic sense. As outlined in section 9.8 of [15], a numerical scheme for a given SDE is considered stochastically stable if, for any finite time interval $[t_0, T]$, there exists a positive constant Δ_0 such that for any $\epsilon > 0$ and any $\delta \in (0, \Delta_0)$, the probability of the discrete time approximation, X_n^δ, starting at X_0^δ and the approximation $\bar{X}_n^\delta$, starting at $\bar{X}_0^\delta$ deviating from each other by more than ϵ over the interval $t_0 \leq t \leq T$ approaches zero as the initial difference between X_0^δ and $\bar{X}_0^\delta$ approaches zero. This we can write compactly as,

$$\lim_{|X_0^\delta - \bar{X}_0^\delta| \to 0} \sup_{t_0 \leq t \leq T} P\left(\left|X_t^\delta - \bar{X}_t^\delta\right| \geq \epsilon\right) = 0.$$

Next, we need a stochastic counterpart to the test equation from the deterministic setting. It is though not possible to define one general test equation because we need to take the assumed noise structure into account. We here define it with an additive noise structure for which we can use the complex Ornstein-Uhlenbeck process,

$$\mathrm{d}X_t = -\mu X_t \mathrm{d}t + \sigma \mathrm{dB}_t, \tag{C.2.25}$$

with $\mathrm{Re}(\mu) > 0$. Let us now suppose that we can write a given scheme, R, with equidistant step size $h \equiv \delta$ applied to the test equations C.2.25 in the recursive form,

$$X_{n+1}^h = X_n^h R(-\mu h) + Z_n^h, \tag{C.2.26}$$

where $Z_0^h, Z_1^h, \ldots$ are random variables which do not depend on the X_0^h, $X_1^h, \ldots$. We can then use the same machinery as for deterministic ODEs to analyze stability. We can hence define the stability region as

$$\mathcal{S} := \{\lambda \in \mathbb{C} : |R(\lambda)| \leq 1\}, \tag{C.2.27}$$

where $\lambda = -h\mu$.
We want to investigate the stability of the additive stochastic Runge Kutta methods by use of this framework. As seen the additive stochastic Runge Kutta methods can be written as,

$$X_{n+1}^h = X_n^h + \lambda \left(\alpha \cdot H^{(0)}\right) + \beta^{(1)} \sigma I_{(1)} + \sigma \beta^{(2)} \frac{I_{(1,0)}}{h},$$

where,

$$H^{(0)} = \left(I - \lambda A^{(0)}\right)^{-1} \left(\hat{X}_n^h + B^{(0)} e \sigma \frac{I_{(1,0)}}{h}\right),$$

and $e = (1, 1, \ldots, 1, 1)^T$. By substitution we receive,

$$\begin{aligned} X_{n+1}^h = X_n^h \left(1 + \lambda \alpha \cdot \left(I - \lambda A^{(0)}\right)^{-1}\right) + \\ \left(I - \lambda A^{(0)}\right)^{-1} B^{(0)} e \sigma \frac{I_{(1,0)}}{h} + \beta^{(1)} \sigma I_{(1)} + \sigma \beta^{(2)} \frac{I_{(1,0)}}{h}. \end{aligned}$$

Since all the terms with an iterated Itô integral are independent of X_n^h and μ, we know from eq. (C.2.26) that $R(\cdot)$ is given by,

$$R(\lambda) = 1 + \lambda \alpha \cdot \left(I - \lambda A^{(0)}\right)^{-1}. \tag{C.2.28}$$

Hence we can plot the stability region for Euler-Maruyama and SRA3 to compare their stability regions.

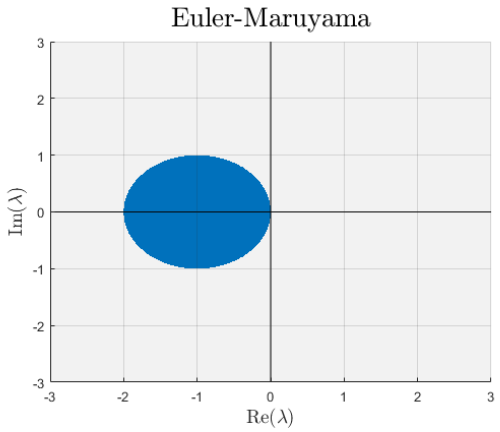


Figure C.5 – Path-wise stable region for the Euler-Maruyama method

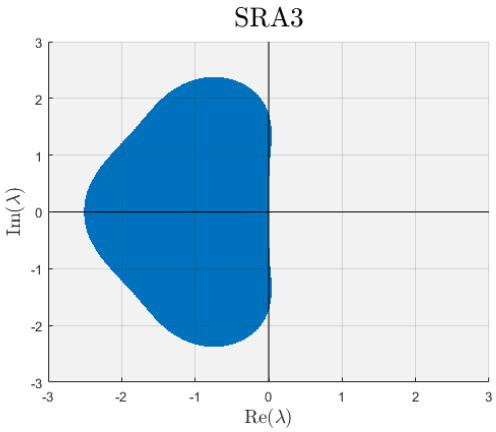


Figure C.6 – Path-wise stable region for the SRA3 method

We see that the SRA3 method is stable for a much larger range than the Euler-Maruyama method.

Moment Stability

Since we are dealing with stochastic differential equations, we not only need to consider stability in a path-wise sense but also in terms of the stability of each moment of the SDE. Therefore, we will examine the mean and mean square stability of our numerical schemes.

Before we do so we have from [16] that for our test equation C.2.25, the gaussian process Z with mean zero and variance $\sigma^2/2Re[\mu]$, is a stationary solution. Moreover, any other solution decays exponentially to this equilibrium solution.

Hence, we will define stability in mean and mean square for a numerical scheme as not diverging from these quantities. Furthermore we would also like to know if the given numerical scheme is biased in the mean and variance w.r.t. the test equation.
In [17] they define,(Definition 2.9.1): The method is said to be asymptotically unbiased with step $h > 0$ if while applying it with the given step to the scalar linear SDE given in eq. (C.2.25) with $\mu > 0$, the distribution of the numerical solution X_n converges as $n \to \infty$ to the normal distribution with zero mean and variance $\sigma^2/2\mu$.

In [18] they suppose that one can write a method applied to eq. (C.2.25) as

$$X_{n+1} = R(\lambda)X_n + Q(\lambda)\sigma\sqrt{h}\xi_n, \tag{C.2.29}$$

where $\lambda = -\mu h$ and $\xi_n \sim \mathcal{N}(0,1)$. We see that this is exactly the same form as eq. (C.2.26) where $Q(\lambda)\sigma\sqrt{h}\xi_n = Z_n^h$. Given the form in eq. (C.2.29) one obtains

$$\mathbb{E}X_{n+1} = R(\lambda)\mathbb{E}X_n, \quad \mathbb{V}X_{n+1} = R^2(\lambda)\mathbb{V}X_n + Q^2(\lambda)\sigma^2 h,$$

and for $|R(\lambda)| < 1$ both variance and expectation will be stable and one has

$$\lim_{n\to\infty} \mathbb{E}X_n = 0, \quad \lim_{n\to\infty} \mathbb{V}X_n = \frac{\sigma^2}{2\mu} \cdot \frac{(-2\lambda)Q^2(\lambda)}{1 - R^2(\lambda)}$$

We hence see that only if $Q^2(\lambda) = \left(1 - R^2(\lambda)\right)/(-2\lambda)$ the numerical scheme will be unbiased. To see the validity of these claims, we try to determine the quantities $\mathbb{E}X_n$ and $\mathbb{V}X_n$

for the Euler Maruyama scheme. For the following analysis, we will assume that $\mu < 0$ means that we just absorb the negative sign in front of the μ in eq. (C.2.25). This will make the analysis easier.

X_{n+1} for the Euler-Maruyama scheme is then given as,

$$\begin{aligned} X_{n+1} &= X_n + \mu\Delta t X_n + \sigma\Delta B_n \\ &= (1+\mu\Delta t)X_n + \sigma\Delta B_n. \end{aligned} \tag{C.2.30}$$

We see that following eq. (C.2.29), $R(\lambda) = (1+\lambda)$, where $\lambda = \mu\Delta t$ and $Q(\lambda) = 1$ as we can write $\Delta B_n = \sqrt{\Delta t}\xi_n$. We start by finding $\mathbb{E}X_{n+1}$.

$$\begin{aligned} \mathbb{E}[X_{n+1}] &= \mathbb{E}[(1+\mu\Delta t)X_n + \sigma\Delta B_n] \\ &= (1+\mu\Delta t)\mathbb{E}[X_n] \\ &= (1+\mu\Delta t)^{n+1}\mathbb{E}[X_0] \\ &= R(\lambda)^{n+1}\mathbb{E}[X_0] \end{aligned} \tag{C.2.31}$$

Thus as expected if $R(\lambda) < 1$ then we have,

$$\lim_{n\to\infty} \mathbb{E}[X_{n+1}] = 0.$$

We proceed with the variance,

$$\begin{aligned} \mathbb{E}[X_{n+1}^2] &= \mathbb{E}[((1+\mu\Delta t)X_n + \sigma\Delta B_n)^2] \\ &= \mathbb{E}[(1+\mu\Delta t)^2 X_n^2] + \mathbb{E}[\sigma^2\Delta B_n^2] + \mathbb{E}[2\sigma\Delta B_n(1+\mu\Delta t)X_n] \\ &= (1+\mu\Delta t)^2\mathbb{E}[X_n^2] + \sigma^2\Delta t. \end{aligned} \tag{C.2.32}$$

We can now perform the recursion we saw used for the expectation and obtain,

$$\begin{aligned} \mathbb{E}\left[X_{n+1}^2\right] &= (1+\mu\Delta t)^2\left((1+\mu\Delta t)^2\mathbb{E}\left[X_{n-1}^2\right] + \sigma^2\Delta t\right) + \sigma^2\Delta t \\ &= (1+\mu\Delta t)^4\mathbb{E}\left[X_{n-1}^2\right] + \left((1+\mu\Delta t)^2 + 1\right)\sigma^2\Delta t \\ &= (1+\mu\Delta t)^{2(n+1)}\mathbb{E}\left[X_0^2\right] + \left((1+\mu\Delta t)^{2n} + \cdots + (1+\mu\Delta t)^2 + 1\right)\sigma^2\Delta t \\ &= (1+\mu\Delta t)^{2(n+1)}\mathbb{E}\left[X_0^2\right] + \frac{(1+\mu\Delta t)^{2(n+1)} - 1}{(1+\mu\Delta t)^2 - 1}\sigma^2\Delta t \\ &= (1+\mu\Delta t)^{2(n+1)}\mathbb{E}\left[X_0^2\right] + \frac{(1+\mu\Delta t)^{2(n+1)} - 1}{2\mu + \mu^2\Delta t}\sigma^2 \\ &= R(\lambda)^{2(n+1)}\mathbb{E}\left[X_0^2\right] + \frac{R(\lambda)^{2(n+1)} - 1}{2\mu + \mu^2\Delta t}\sigma^2. \end{aligned} \tag{C.2.33}$$

Hence we again have that if $R(\lambda) < 1$ then,

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}[X_{n+1}^2] &= -\frac{\sigma^2}{2\mu + \mu^2\Delta t} \\ &= -\frac{\sigma^2}{2\mu} + \frac{\sigma^2\Delta t}{2\Delta t\mu + 4}. \end{aligned} \tag{C.2.34}$$

We notice that if we go back to the setting where $\mu > 0$ but entered into eq. (C.2.25) as $-\mu$ then

$$\lim_{n\to\infty} \mathbb{E}[X_{n+1}^2] = \frac{\sigma^2}{2\mu} + \frac{\sigma^2 \Delta t}{-2\Delta t\mu + 4}. \tag{C.2.35}$$

Hence when Δt goes to zero, the variance of the Euler-Maruyama tends toward an unbiased estimator. We further check if $Q^2(\lambda) = \left(1 - R^2(\lambda)\right)/(-2\lambda)$ also in the limit where $\Delta t \to 0$.

$$\begin{aligned}
\lim_{\Delta t\to 0} \left(1 - R^2(\lambda)\right)/(-2\lambda) &= \lim_{\Delta t\to 0} -\frac{1-(1+\lambda)^2}{2\lambda} \\
&= \lim_{\Delta t\to 0} \frac{\lambda}{2} + 1 \\
&= \lim_{\Delta t\to 0} \frac{\mu\Delta t}{2} + 1 \\
&= 1
\end{aligned}$$

Hence as expected and we will now proceed to the SRA3 method.

The SRA3 method: For the SRA3 method, we use $\mu > 0$ and the form for the test equation given in eq. (C.2.25). Hence we have the following quantities that have

$$R(\lambda) = -\frac{1}{6}\lambda^3 + \frac{1}{2}\lambda^2 - \lambda + 1, \qquad Q(\lambda) = \frac{\sqrt{3\lambda^2 - 9\lambda + 9}}{3}$$

The expectation is trivial and will be both unbiased and stable as long $|R(\lambda)| < 1$, i.e. as long we have path-wise stability the mean will also be stable and unbiased.

The variance is now given as

$$\begin{aligned}
\mathbb{V}X_n &= \frac{\sigma^2}{2\mu} \cdot \frac{(-2\lambda)Q^2(\lambda)}{1 - R^2(\lambda)} \\
&= \frac{\sigma^2 \Delta t \left(\frac{(\Delta t\mu - 2)^2}{4} + \frac{\Delta t^2\mu^2}{12}\right)}{1 - \frac{(\Delta t^3\mu^3 - 3\Delta t^2\mu^2 + 6\Delta t\mu - 6)^2}{36}} \\
&= \frac{\sigma^2}{2\mu} - \frac{\sigma^2 \Delta t^2 \mu \left(\Delta t^3\mu^3 - 6\Delta t^2\mu^2 + 21\Delta t\mu - 24\right)}{2\Delta t^5\mu^5 - 12\Delta t^4\mu^4 + 42\Delta t^3\mu^3 - 96\Delta t^2\mu^2 + 144\Delta t\mu - 144}
\end{aligned}$$

We see that the bias term for the variance when Δt is not zero is much more complicated for the SRA3 method. We will plot it together with the Euler-Maruyama method to understand it better. In fig. C.7 we see the bias for the two methods with $\sigma = \mu = 2$. We see that the SRA3 method reduces the bias in variance twice as fast in the loglog domain compared to the Euler-Maruyama method.

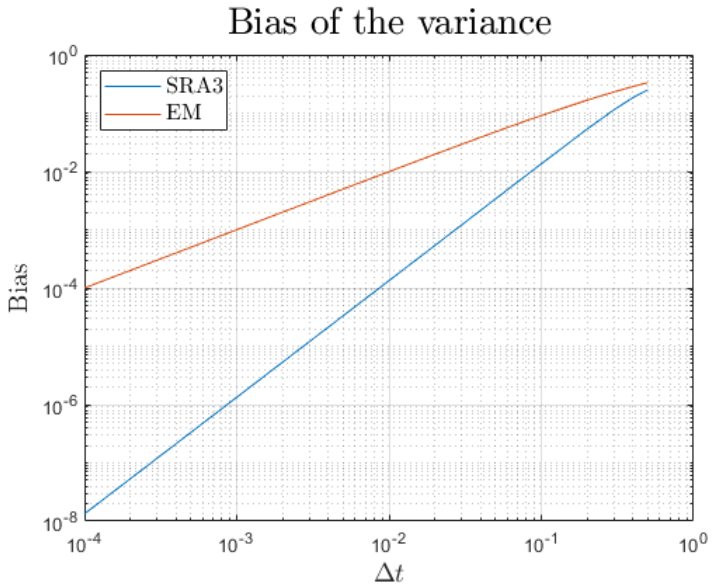


Figure C.7 – Bias in the variance shown for the SRA3 and Euler-Maruyama method when $\mu = \sigma = 2$

C.2.5 The Lamperti Transform

It is possible to perform a stability analysis for the methods with scalar noise structure, however, we will not delve into this as it is more complex, and in the next section, we will only be utilizing methods with additive noise. Instead, we will present a technique known as the Lamperti transform, which can convert the noise structure to an additive structure. This transformation is applicable for SDEs with scalar noise structures and some multivariate noise structures as stated in theorem 4 of [19]:

Let $\{\boldsymbol{X}_t : t \geq 0\}$ be an Itô diffusion given by,

$$d\boldsymbol{X}_t = \boldsymbol{f}\left(\boldsymbol{X}_t, t\right) dt + \boldsymbol{\sigma}\left(\boldsymbol{X}_t, t\right) \boldsymbol{R}(t) d\boldsymbol{B}_t,$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n\times n}$ is any matrix function of t, and $\boldsymbol{\sigma}\left(\boldsymbol{X}_t, t\right) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal elements $\sigma_{i,i}\left(\boldsymbol{X}_t, t\right)$ given by,

$$\sigma_{i,i}\left(\boldsymbol{X}_t, t\right) = \sigma_i\left(X_{i,t}, t\right).$$

Then the transformation,

$$Z_{i,t} = \psi_i\left(X_{i,t}, t\right) = \int \frac{1}{\sigma_i(x,t)} dx \bigg|_{x=X_{i,t}},$$

will result in an Itô process with the i'th element given by,

$$dZ_{i,t} = \left(\frac{\partial}{\partial t}\psi_i(x,t)\bigg|_{x=\psi^{-1}(Z_{i,t},t)} + \frac{f_i\left(\psi^{-1}\left(Z_t,t\right),t\right)}{\sigma_i\left(\psi_i^{-1}\left(Z_{i,t},t\right),t\right)} - \frac{1}{2}\frac{\partial}{\partial x}\sigma_i\left(\psi_i^{-1}\left(Z_{i,t},t\right),t\right)\right) dt + \sum_{j=1}^{n} r_{ij}(t) dB_{j,t},$$

where $r_{ij}(t)$ are elements of $\boldsymbol{R}(t)$ and,

$$\boldsymbol{X}_t = \boldsymbol{\psi}^{-1}\left(\boldsymbol{Z}_t, t\right).$$

In the case of mixed multiplicative and additive noise we have an SDE of the form,

$$dX_t = f\left(t, X_t\right) dt + \left(\sigma_M X_t + \sigma_A\right) dB_t,$$

with $\sigma_M > 0$ and $\sigma_A > 0$, the transform becomes element-wise in the system. In [20] they consider the one-dimensional case. Since $\psi\left(t, X_t\right) = \int \left(\frac{1}{\sigma_M X_t + \sigma_A}\right) dx \Big|_{x=X_t} = \frac{\log(\sigma_M X_t + \sigma_A)}{\sigma_M}$, then $X_t = \frac{\exp(\sigma_M Z_t) - \sigma_A}{\sigma_M}$ and,

$$\begin{aligned} dZ_t &= \tilde{f}\left(t, X_t\right) dt + dB_t \\ \tilde{f}\left(t, X_t\right) &= \frac{f\left(t, X_t\right)}{\sigma_M X_t + \sigma_A} - \frac{1}{2}\sigma_M. \end{aligned}$$

The stochastic Runge-Kutta method with additive noise that was previously introduced, can be transformed in a way to handle scalar noise as presented in [20]. We change the algorithm as follows,

$$X_{n+1} = \psi^{-1}\left(\psi\left(X_n\right) + \sum_{i=1}^{s} \alpha_i \tilde{f}\left(t_n + c_i^{(0)} h, H_i^{(0)}\right) + \sum_{i=1}^{s} \left(\beta_i^{(1)} I_{(1)} + \beta_i^{(2)} \frac{I_{(1,0)}}{h}\right) \tilde{g}\left(t_n + c_i^{(1)} h\right)\right),$$

with stages,

$$H_i^{(0)} = \psi\left(X_n\right) + \sum_{j=1}^{s} A_{ij}^{(0)} \tilde{f}\left(t_n + c_j^{(0)} h, H_j^{(0)}\right) h + \sum_{j=1}^{s} B_{ij}^{(0)} \tilde{g}\left(t_n + c_j^{(1)} h\right) \frac{I_{(1,0)}}{h},$$

where ψ is the element-wise function,

$$\psi_i(x) = \begin{cases} \frac{\log\left(\sigma_{i,M} x + \sigma_{i,A}\right)}{\sigma_{i,M}} & \sigma_{i,M} > 0, \sigma_{i,A} > 0 \\ \frac{\log(x)}{\sigma_{i,M}} & \sigma_{i,M} > 0, \sigma_{i,A} = 0 \\ x & \text{otherwise} \end{cases}$$

$$\psi_i^{-1}(z) = \begin{cases} \frac{\exp\left(\sigma_{i,M} z\right)}{\sigma_{i,M}} & \sigma_{i,M} > 0, \sigma_{i,A} > 0 \\ \frac{\exp(z)^D}{\sigma_{i,M}} & \sigma_{i,M} > 0, \sigma_{i,A} = 0 \\ x & \text{otherwise} \end{cases}$$

and,

$$\tilde{g}_i(t) = \begin{cases} 1 & \sigma_{i,M} > 0 \\ \sigma_{i,A} & \text{otherwise} \end{cases}$$

This can be summarized as performing all internal operations in Z-space (where the equation is additive) but saving each step in X-space.

D | Parameter Estimation

In the following, we will cover some algorithms to do parameter estimations for SDEs. As a stylus example of each method, we will use the Ornstein Uhlenbeck process that is a linear time-invariant SDE. Linear time-invariant SDEs can in general be written as,

$$\mathrm{d}x(t) = \mathbf{F}x(t)\mathrm{d}t + \mathbf{G}\mathrm{dB}(t) \tag{D.0.1}$$

,
where $\mathbf{F}$, $\mathbf{G}$ are constants. We will use this general form first and then translate it into the Ornstein-Uhlenbeck process we have seen in earlier chapters.

D.1 Exact Maximum Likelihood

Exact parameter estimation in SDEs becomes very challenging and even impossible very fast. For the Ornstein Uhlenbeck process though it is possible and in this section, we will derive the exact maximum likelihood for this process. This also creates a reference for all the subsequent methods. In [10], they derive the expectation and covariance equations of linear time-invariant SDE, eq. (D.0.1), by use of the Kolmogorov forward equation,

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \mathbf{F}\mathbf{m}, \tag{D.1.1}$$

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}t} = \mathbf{F}\Sigma + \Sigma\mathbf{F}^\top + \mathbf{G}\mathbf{G}^\top, \tag{D.1.2}$$

with the following solutions,

$$\mathbf{m}(t) = \exp\left(\mathbf{F}\left(t - t_0\right)\right)\mathbf{m}\left(t_0\right) \tag{D.1.3}$$

$$\begin{aligned}\Sigma(t) =& \exp\left(\mathbf{F}\left(t - t_0\right)\right)\Sigma\left(t_0\right)\exp\left(\mathbf{F}\left(t - t_0\right)\right)^\top \\ &+ \int_{t_0}^{t} \exp(\mathbf{F}(t-\tau))\mathbf{G}\mathbf{G}^\top \exp(\mathbf{F}(t-\tau))^\top \mathrm{d}\tau.\end{aligned} \tag{D.1.4}$$

The transition density is then given as,

$$p(\mathbf{x}(t_{k+1}) \mid \mathbf{x}(t_k)) = \mathcal{N}(\mathbf{x}(t_{k+1}) \mid \mathbf{m}(t_{k+1} \mid t_k), \Sigma(t_{k+1} \mid t_k)), \tag{D.1.5}$$

where,

$$\mathbf{m}(t_{k+1} \mid t_k) = \exp(\mathbf{F}(t_{k+1} - t_k))\mathbf{x}(t_k) \tag{D.1.6}$$

$$\Sigma(t_{k+1} \mid t_k) = \int_{t_k}^{t_{k+1}} \exp(\mathbf{F}(t_{k+1} - \tau))\mathbf{G}\mathbf{G}^\top \exp(\mathbf{F}(t_{k+1} - \tau))^\top \mathrm{d}\tau. \tag{D.1.7}$$

We see that matrix exponentials are included in both eq. (D.1.6) and eq. (D.1.7). This can be a computational issue, so to avoid dealing with that we will only derive explicit formulas for a scalar case here. For the scalar case we can convert the linear time-invariant SDE into the Ornstein Uhlenbeck process by assuming $\mathbf{f} < 0$ and $\mathbf{g} > 0$ or adding a minus in front of $\mathbf{f}$ in eq. (D.0.1) and assuming $\mathbf{f} > 0$. We will continue using the last form previously mentioned, as it has been our approach thus far and instead of writing $\mathbf{f}$ and $\mathbf{g}$ we will write μ and σ. This will allow us to derive the mean, eq. (D.1.6), and variance, eq. (D.1.7), which takes the following forms,

$$\mathbf{m}(t_{k+1} \mid t_k) = \exp(-\mu\Delta t)x(t_k) \tag{D.1.8}$$

$$\Sigma(t_{k+1} \mid t_k) = \frac{1}{2\mu}\sigma^2\left(1 - \exp(-2\mu\Delta t)\right), \tag{D.1.9}$$

where $\Delta t = t_{k+1} - t_k$. We can now derive the negative log-likelihood to be

$$\ell(\mathbf{m}, \Sigma) = \sum_{k=0}^{T-1}\left[\frac{1}{2}\log(2\pi\Sigma(t_{k+1} \mid t_k)) + \frac{1}{2\Sigma(t_{k+1} \mid t_k)}\left(x\left(t_{k+1}\right) - \mathbf{m}(t_{k+1} \mid t_k)\right)^2\right] \tag{D.1.10}$$

We differentiate eq. (D.1.10) w.r.t. the parameters μ and σ and obtain the estimators

$$\mu_{\mathrm{ML}} = -\frac{1}{\Delta t}\log\left[\frac{\sum_{k=0}^{T-1} x\left(t_k\right) x\left(t_{k+1}\right)}{\sum_{k=0}^{T-1} x\left(t_k\right) x\left(t_k\right)}\right] \tag{D.1.11}$$

$$\sigma_{\mathrm{ML}} = \sqrt{\frac{1}{T}\left(\frac{2\mu_{\mathrm{ML}}}{1 - \exp\left(-2\mu_{\mathrm{ML}}\Delta t\right)}\right)\sum_{k=0}^{T-1}\left(x\left(t_{k+1}\right) - \exp\left(-\mu_{\mathrm{ML}}\Delta t\right) x\left(t_k\right)\right)^2}. \tag{D.1.12}$$

D.2 Pseudo-likelihood Methods

The estimates in the previous section are only possible to derive for a very small subset of SDEs because they rely on an analytical solution to the forward Kolmogorov equation. Hence we would like a more general method that could estimate parameters for an arbitrary SDE. To obtain such a family of methods we can use the numerical approximation schemes developed in the previous chapter. These approximation schemes do not approximate the transition density directly but the path of the process. We can hence instead use the transition density of the scheme instead of the true transition density. Such methods are called pseudo-likelihood methods.

D.2.1 The Euler Approximation

The simplest method in the family of pseudo-likelihood methods is obtained by approximating the sample path with the Euler-Maruyama method. If we have a general SDE given by

$$\mathrm{d}x(t) = \mathbf{f}(x(t), t; \theta)\mathrm{d}t + \mathbf{g}(x(t), t; \theta)\mathrm{dB}(t) \tag{D.2.1}$$

We can now discretize eq. (D.2.1) using the Euler-Maruyama scheme and obtain

$$\hat{\mathbf{x}}\left(t_{k+1}\right) = \hat{\mathbf{x}}\left(t_k\right) + \mathbf{f}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right)\Delta t + \mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right)\Delta \mathrm{B}_k, \tag{D.2.2}$$

where $\Delta \mathrm{B}_k \sim \mathcal{N}(0, \Delta t)$. We know from [21] that the transition density of the Euler-Maruyama scheme is given by a Gaussian, for which we can derive the expectation and variance.

$$\begin{aligned}\mathbb{E}^{EM}\left[\hat{\mathbf{x}}\left(t_{k+1}\right) \mid \hat{\mathbf{x}}\left(t_k\right), \theta\right] =& \mathbb{E}\left[\hat{\mathbf{x}}\left(t_k\right) + \mathbf{f}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta t + \mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta \mathrm{B}_k\right] \\ =& \mathbb{E}\left[\hat{\mathbf{x}}\left(t_k\right)\right] + \mathbb{E}\left[\mathbf{f}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta t\right] + \mathbb{E}\left[\mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta \mathrm{B}_k\right] \\ =& \hat{\mathbf{x}}\left(t_k\right) + \mathbf{f}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta t\end{aligned} \tag{D.2.3}$$

$$\begin{aligned}\mathbb{V}^{EM}\left[\hat{\mathbf{x}}\left(t_{k+1}\right) \mid \hat{\mathbf{x}}\left(t_k\right), \theta\right] =& \mathbb{V}\left[\hat{\mathbf{x}}\left(t_k\right) + \mathbf{f}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta t + \mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta \mathrm{B}_k\right] \\ =& \mathbb{V}\left[\hat{\mathbf{x}}\left(t_k\right)\right] + \mathbb{V}\left[\mathbf{f}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta t\right] + \mathbb{V}\left[\mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right) \Delta \mathrm{B}_k\right] \\ =& 0 + 0 + \mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right)^2 \mathbb{V}\left[\Delta \mathrm{B}_k\right] \\ =& \mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right)^2 \Delta t\end{aligned} \tag{D.2.4}$$

We hence obtain the approximated transition density

$$\begin{aligned}&\hat{p}^{EM}\left(\mathbf{x}\left(t_{k+1}\right) \mid \mathbf{x}\left(t_k\right), \theta\right) = \\ &\qquad \mathcal{N}(\mathbf{x}\left(t_{k+1}\right) \mid \mathbf{x}\left(t_k\right) + \mathbf{f}\left(\mathbf{x}\left(t_k\right), t_k; \theta\right) \Delta t, \mathbf{g}\left(\hat{\mathbf{x}}\left(t_k\right), t_k; \theta\right)^2 \Delta t),\end{aligned} \tag{D.2.5}$$

which results in the following negative log-likelihood

$$\hat{\ell}^{EM}(\theta) = -\sum_{k=0}^{T-1} \log \hat{p}^{EM}\left(\mathbf{x}\left(t_{k+1}\right) \mid \mathbf{x}\left(t_k\right), \theta\right). \tag{D.2.6}$$

The effect of Δt on the Euler approximation

Before we move to more advanced methods we will investigate which problems can emerge from using methods from the family of pseudo-likelihood methods. If Δt is not small enough, the estimates will include bias even for the very simple Ornstein-Uhlenbeck model. In [21] they give the Euler's pseudo transition density and the true transition density for the Ornstein-Uhlenbeck model which both are Gaussian distributed. The mean and variance for the true transition density are given by,

$$\mathbb{E}[x|\Delta t] = xe^{-\mu \Delta t}, \quad \mathbb{V}[x|\Delta t] = \frac{\sigma^2\left(1 - e^{-2\mu \Delta t}\right)}{2\mu},$$

and for the Euler approximation we have

$$\mathbb{E}^{\text{Euler}}\left[x|\Delta t\right] = x\left(1 - \mu \Delta t\right), \quad \mathbb{V}^{\text{Euler}}\left[x|\Delta t\right] = \sigma^2 \Delta t,$$

We see that $\mathbb{E}^{\text{Euler}}$ and $\mathbb{V}^{\text{Euler}}$ only coincide with the true values, if $\Delta t \to 0$. In [21] they show that for an SDE like the geometric Brownian motion, which does not have a transition density that is Gaussian, the Euler approximation is not even consistent. To solve the problems related to the Euler approximation we can do two things. The first fix is to use a more precise scheme, like the Milstein or SRA3 scheme. We will in the next section derive and implement the pseudo-likelihood for the Ornstein-Uhlenbeck model using the SRA3 method. This gives us a method that works for much smaller Δt than the Euler approximation but does still not give us a method that can estimate the parameters for an arbitrary Δt and we still have the problem of inconsistency.
The second option we will present in this report is simulated maximum likelihood. This method works for an arbitrary Δt and also solves the problem of inconsistency but comes at the cost of a lot of computations. We will not implement it but briefly explain it in section D.3.

D.2.2 Estimation with SRA3

We now derive the pseudo-likelihood for the Ornstein-Uhlenbeck model using the SRA3 method. First we derive the distribution for the two iterated Itô integrals that enter into the SRA3 method. They are given by,

$$\begin{aligned} I_{(1)} &= \mathrm{dB}(t) \sim \mathcal{N}(0, \Delta t) \\ I_{(1,0)} &= \frac{1}{2}\Delta t I_{(1)} - \frac{1}{2\sqrt{3}}\Delta t \xi, \end{aligned} \tag{D.2.7}$$

where $\xi \sim \mathcal{N}(0, \Delta t)$. The expectation for both terms are easily determined to be 0. Next we need to determine their variances,

$$\mathbb{V}\left[I_{(1)}\right] = \Delta t \tag{D.2.8}$$

$$\begin{aligned} \mathbb{V}\left[I_{(1,0)}\right] &= \mathbb{V}\left[\frac{1}{2}\Delta t I_{(1)} - \frac{1}{2\sqrt{3}}\Delta t \xi\right] \\ &= \mathbb{V}\left[\frac{1}{2}\Delta t I_{(1)}\right] + \mathbb{V}\left[\frac{1}{2\sqrt{3}}\Delta t \xi\right] \\ &= \frac{1}{4}\Delta t^2 \mathbb{V}\left[I_{(1)}\right] + \frac{1}{12}\Delta t^2 \mathbb{V}\left[\xi\right] \\ &= \frac{1}{3}\Delta t^3 \end{aligned} \tag{D.2.9}$$

Lastly, we need the covariance where we use that dB and ξ are independent,

$$\begin{aligned} \mathbf{Cov}[I_{(1)}, I_{(1,0)}] &= \mathbb{E}[I_{(1)} I_{(1,0)}] - \mathbb{E}[I_{(1)}]\mathbb{E}[I_{(1,0)}] \\ &= \mathbb{E}[I_{(1)} I_{(1,0)}] \\ &= \mathbb{E}[\frac{1}{2}\Delta t \mathrm{dB}^2 - \frac{1}{2\sqrt{3}}\Delta t \mathrm{dB}\xi] \\ &= \frac{1}{2}\Delta t \mathbb{E}[\mathrm{dB}^2] - \frac{1}{2\sqrt{3}}\Delta t \mathbb{E}[\mathrm{dB}\xi] \\ &= \frac{1}{2}\Delta t \mathbb{E}[\mathrm{dB}^2] - \frac{1}{2\sqrt{3}}\Delta t \mathbb{E}[\mathrm{dB}]\mathbb{E}[\xi] \\ &= \frac{1}{2}\Delta t^2. \end{aligned} \tag{D.2.10}$$

For the Ornstein Uhlenbeck model, we can now determine the transition density when we use the SRA3 method. We know that $f(x,t) = -\mu x$ and $g(x,t) = \sigma$ which means the increments are then given by

$$\hat{\mathbf{x}}\left(t_{k+1}\right) = \frac{\left(-\Delta t^3 \mu^3 + 3\Delta t^2 \mu^2 - 6\Delta t \mu + 6\right)\hat{\mathbf{x}}\left(t_k\right)}{6} + \sigma(I_{(1)} - \mu I_{(1,0)}) \tag{D.2.11}$$

For the expectation and variance we get

$$\mathbb{E}_{OU}^{SRA3}\left[\hat{\mathbf{x}}\left(t_{k+1}\right) \mid \hat{\mathbf{x}}\left(t_k\right), \mu, \sigma^2\right] = \frac{\left(-\Delta t^3 \mu^3 + 3\Delta t^2 \mu^2 - 6\Delta t \mu + 6\right)\hat{\mathbf{x}}\left(t_k\right)}{6} \tag{D.2.12}$$

$$\begin{aligned} \mathbb{V}_{OU}^{SRA3}\left[\hat{\mathbf{x}}\left(t_{k+1}\right) \mid \hat{\mathbf{x}}\left(t_k\right), \mu, \sigma^2\right] =& \mathbb{V}[\sigma I_{(1)} - \sigma\mu I_{(1,0)}] \\ =& \sigma^2 \mathbb{V}[I_{(1)}] + \sigma^2\mu^2 \mathbb{V}[I_{(1,0)}] - 2\sigma^2\mu \mathbf{Cov}[I_{(1)}, I_{(1,0)}] \\ =& \sigma^2(\Delta t - \mu\Delta t^2 + \mu^2 \frac{\Delta t^3}{3}) \end{aligned}$$

We hence obtain the approximated transition density

$$\hat{p}_{OU}^{SRA3}\left(\mathbf{x}\left(t_{k+1}\right) \mid \mathbf{x}\left(t_k\right), \mu, \sigma^2\right) =$$
$$\mathcal{N}(\mathbf{x}\left(t_{k+1}\right) \Bigg| \frac{\left(-\Delta t^3\mu^3 + 3\Delta t^2\mu^2 - 6\Delta t\mu + 6\right)\hat{\mathbf{x}}\left(t_k\right)}{6},\ \sigma^2\left[\Delta t - \mu\Delta t^2 + \mu^2\frac{\Delta t^3}{3}\right])$$

The negative log-likelihood is now given as

$$\hat{\ell}_{OU}^{SRA3}(\mu, \sigma^2) = -\sum_{k=0}^{T-1} \log \hat{p}_{OU}^{SRA3}\left(\mathbf{x}\left(t_{k+1}\right) \mid \mathbf{x}\left(t_k\right), \mu, \sigma^2\right). \tag{D.2.13}$$

The SRA3 method is highly accurate with a strong order of 1.5 and a weak order of 3.0, that is three times as accurate as the Euler-Maruyama method. However, as discussed earlier, it has the limitation of only being applicable to additive noise. Despite this limitation, it is still suitable for all univariate processes and in some cases for multivariate processes by use of the Lamperti transformation, as discussed in section C.2.5.

D.3 Simulated Maximum Likelihood

The simulated maximum likelihood method is not part of the family of pseudo-likelihood methods but part of another family called approximated likelihood methods. These methods do not try to approximate the paths of a diffusion but instead provide direct a approximation of the transition density. The following is based on section 3.3.2 in [21].

To derive the simulated maximum likelihood we let $p_\theta(\Delta t, y \mid x)$ be the true transition density of $X_{t+\Delta t}$ at point y given $X_t = x$. When the time step Δt is too large, we know that our pseudo-likelihood methods will eventually break down and give poor estimates of $p_\theta(\Delta, y \mid x)$. To solve this problem we now consider a smaller $\delta t << \Delta t$, for example $\delta t = \Delta t/N$ for N large enough. Then we can use the Chapman-Kolmogorov equation as follows,

$$p_\theta(\Delta t, y \mid x) = \int p_\theta(\delta t, y \mid z) p_\theta(\Delta t - \delta t, z \mid x)\mathrm{d}z = \mathbb{E}_z\left\{p_\theta(\delta t, y \mid z) \mid \Delta t - \delta t\right\},$$

which means that $p_\theta(\Delta, y \mid x)$ is seen as the expected value over all possible transitions of the process from time $t + (\Delta t - \delta t)$ to $t + \Delta t$, taking into account that the process was in x at time t. To realize this, it is necessary to simulate trajectories of the process X from t to $t + (\Delta t - \delta t)$ using some numerical scheme as the ones presented in chapter C. Further we also assume δt is chosen small enough, for the given scheme to be accurate. If the Euler-Maruyama scheme is chosen, then $p_\theta(\Delta t - \delta t, z \mid x)$ is nothing but the Euler-Maruyama transition density, that is Gaussian. It is now necessary to simulate M trajectories of the process in order to estimate $p_\theta(\Delta t, y \mid x)$ by use of Monte Carlo. Let $\phi_\theta(\delta t, y \mid z)$ be the Euler-Maruyama transition density given in eq. (D.2.5). To calculate an estimate of $p_\theta(\Delta, y \mid x)$, we need to simulate M trajectories starting at $X_0 = x$ using a numerical scheme, e.g. the Euler-Maruyama method, with time step δt up to time $\Delta t - \delta t$. We denote the last value of the m-th trajectory as $z_m, m = 1, \ldots, M$. Finally, we obtain the Monte Carlo estimate as,

$$\hat{p}_\theta^{(N,M)}\left(\Delta t, X_{t+\Delta t} \mid x\right) = \frac{1}{M}\sum_{i=1}^{M} \phi_\theta\left(\delta t, X_{t+\Delta t} \mid z_i\right).$$

In principle, by increasing the number of simulated trajectories M, one can obtain any degree of accuracy in the estimation provided that the process is regular enough and δ is also small enough. We see that this method works with an arbitrary Δt and in [22] they showed that when N and M go to infinity the approximated transition density converges to the true density and hence it is also consistent. This though comes with a price of huge computational demands because we have to simulate paths $N \cdot M$ times. We will not implement the method here but refer to [21] or the software SDE toolbox.

D.4 Numerical experiments

To compare the exact likelihood and the Euler and SRA3 approximation, we will simulate 200 Ornstein-Uhlenbeck processes with $\mu = 2$, $\sigma^2 = 2$, $X_0 = 5$, $T = 200$ and vary the step size, Δt, in the set, $\{0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1.0\}$. We then estimate μ and σ^2 using the derived methods. We see in fig. D.1 and fig. D.2 that the Euler approximation gives reliable estimates up to a step size of 0.001 where it starts to become unstable. The SRA3 approximation gives reliable results until a much higher step size. We see that for μ it follows the closed-form solution until 0.1 and for σ it follows the closed-form solution until 0.01.

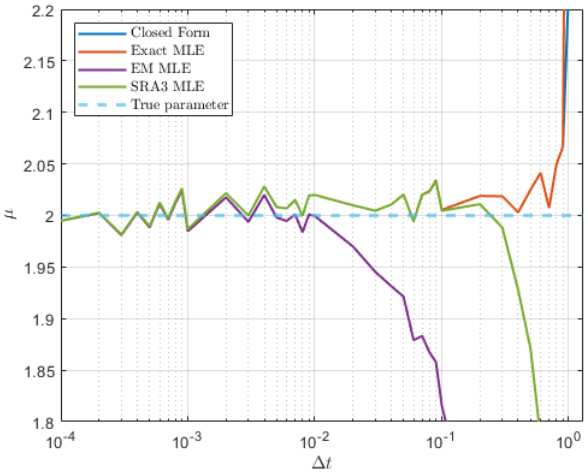


Figure D.1

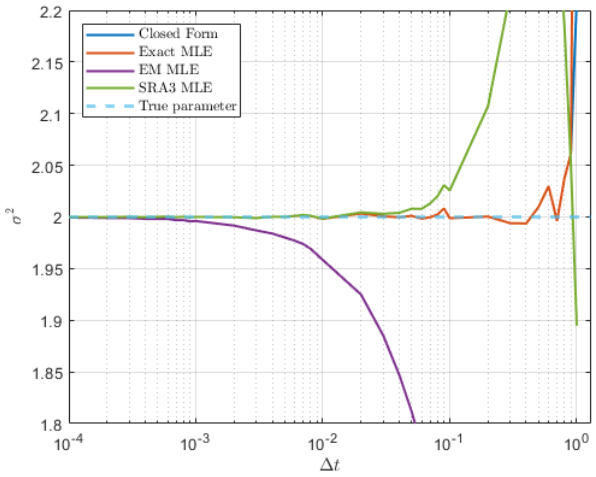


Figure D.2

Hence we will proceed only with the SRA3 approximation. One problem we can face with all the approximations is inaccurate results due to a too short time series. We will now investigate how the different values, T, σ^2 and μ, can make the estimates bad.

D.4.1 Vary T, fix μ and σ

We first fix $\mu = 2$ and $\sigma^2 = 4$ and then we vary T in the set $\{10, 50, 250, 500\}$.

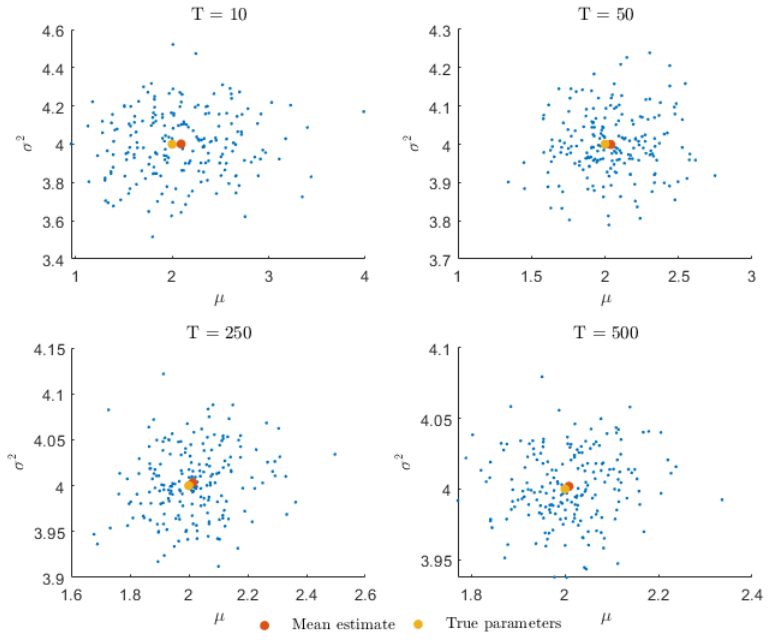


Figure D.3

As shown in fig. D.3, when the value of T is large, the variation in individual estimates is reduced, and the precision of the mean estimate is closer to the true parameter value than when T is low. Additionally, the figure indicates that there is a consistent bias to the right

of the true parameter value when estimating μ. To further investigate this, all simulations were plotted in a single graph.

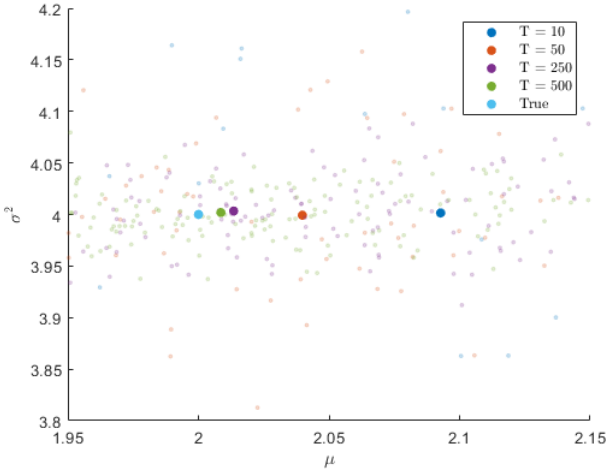


Figure D.4

We see in fig. D.4 that the higher the value of T is the lower the bias to the right is.

D.4.2 Vary σ, fix μ and T

We now try to fix $\mu = 2$ and $T = 20$ and then we vary σ^2 in the set $\{0.1, 1.0, 2.5, 5.0\}$. We see in fig. D.5 that again a bias to the right of the true μ is present. For large values of σ^2 we see a larger bias to the right than for small values of σ^2.

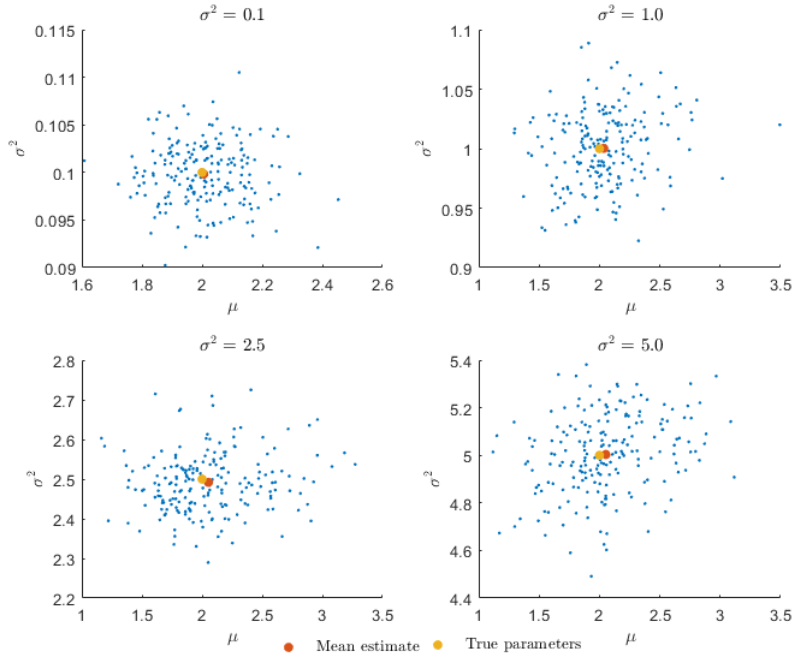


Figure D.5

In fig. D.6 we have plotted in one figure and see that for high values of σ^2 the estimate of μ is biased.

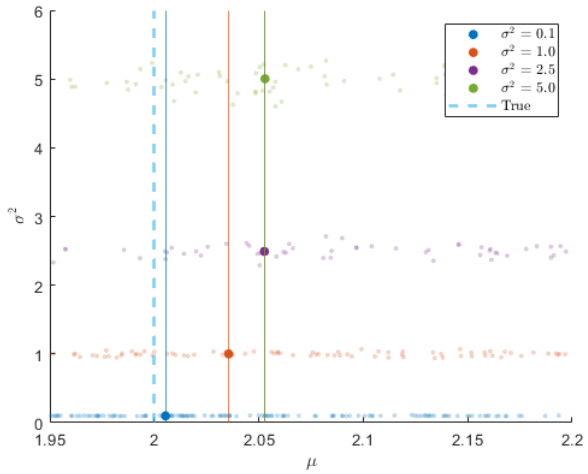


Figure D.6

D.4.3 Vary μ, fix σ and T

Lastly we try to fix $\sigma^2 = 4$ and $T = 20$ and then we vary μ in the set $\{0.1, 1.0, 4.0, 10.0\}$.

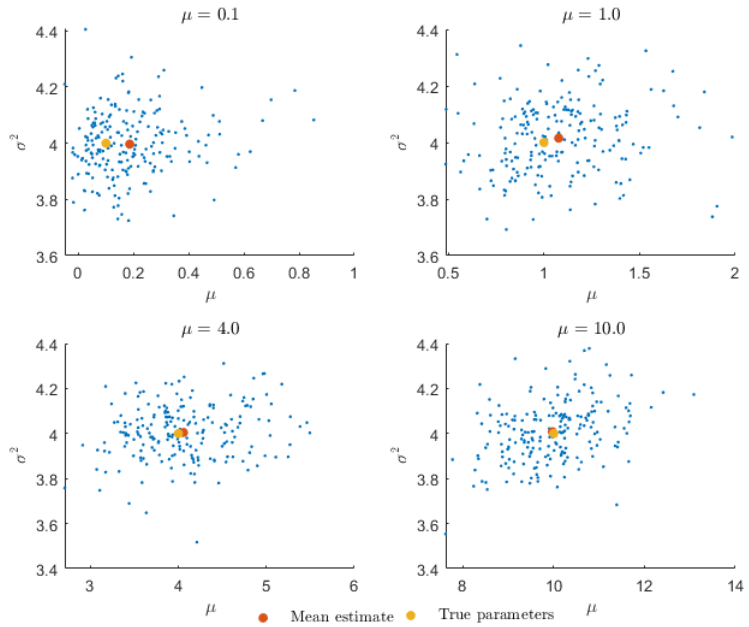


Figure D.7

We see in fig. D.7 that again a bias to the right for estimates of μ is present. Now it is for small values of μ we see the bias. We try to plot in one figure. To do this we normalize all values of μ to 1.

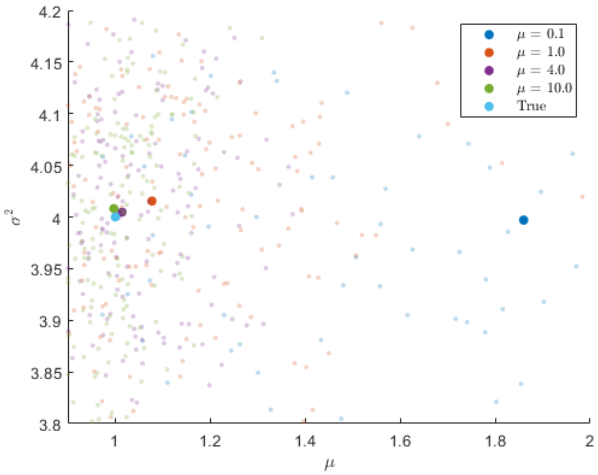


Figure D.8

In fig. D.8 we see the same picture. Hence we can conclude that if one has a very short time series and gets estimates which show a small μ and a large σ^2 one should be careful.

E | Application

In the following, we will use all the introduced methods on the German power prices. We will use the German EPEX Spot prices and EEX monthly power futures prices. We will consider two different periods; one for the entire dataset, and one for only the first years. In the following, we will cover the analysis done on each data set:

- **The entire dataset:** this covers the period from 2002 to 2009 which they also analyze in [2]. We will consider the same period to ensure that we can compare our analysis and that we see the same signals as in [2]. However, the entire data section is affected by macroeconomics that our model will never be able to describe and hence the modelling of the entire dataset has been moved to appendix E.
- **The short period:** We will limit our analysis to a shorter time frame that is not significantly impacted by macroeconomic, political, and energy market factors that our model and inputs cannot account for. Using the long data series discussed in [2] would lead to systematic variance and violate assumptions of independence. To ensure we have enough data, we will use a time period from 2002 to December 1st, 2004 for the spot price and all future price series.

In the following, we will cover some historical notes that can be seen as movements in the data series. Then we will cover some general statistical properties of the spot prices.

E.1 Data Introduction

In this section, we will provide a brief overview of the available data and explain how the series is created. The purpose of this section is to increase understanding and acknowledge that our model, which is based solely on price signals, will not fully capture all aspects of the data due to the lack of other signals.

E.1.1 Spot Prices

The spot prices are German EPEX spot prices. These are hourly day-ahead prices, however, we constructed daily averages as in [2] to be able to compare results. This also simplifies our modeling as the intra-day patterns need not also be described by our model.

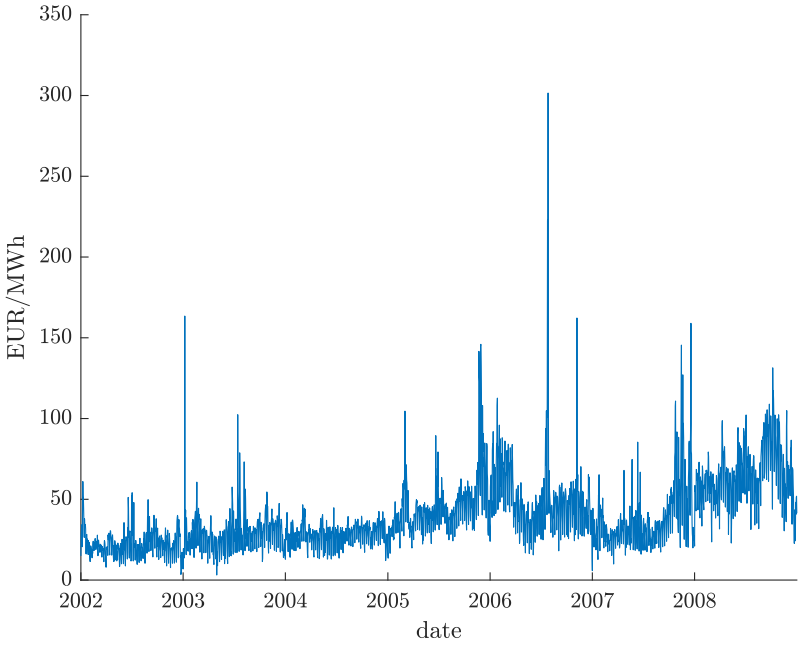


Figure E.1 – The EPEX spot price over the considered time period.

In fig. E.1, we note that the series starts somewhere around 20 EUR/MWh and ends somewhere closer to 50 EUR/MWh. There are several spikes in the signal and these are mostly in a positive direction. In the following, we cover a couple of characteristics of the price series in the period. All the historical notes and views are from [1].

- **The price is low in the beginning**. The liberalized spot market opened in 2002. Prior to that, the energy grids were controlled by regional monopolies with integrated utilities that offered and owned all services. The consumers could not freely choose utilities and the market was by definition not competitive. When the market was liberalized, there was an overcapacity of generating assets which kept prices low. The low prices meant that many old, depreciated power plants were shut down in early 2000.
- **CO_2 prices were introduced in 2005**. This led to an increase in price and accelerated the obsolescence of old, inefficient power plants.
- **Rising natural gas prices around 2006**. The large increase in power prices prior to and in 2006 is caused by a great increase in natural gas prices and in turn also in the spot prices.
- **Excessive CO_2 certificates in 2007**. The initial EU CO_2 scheme issued too many certificates which meant that the price of CO_2 dropped to zero throughout 2007 which drove down power prices.
- **A steady increase in prices until the end of 2008**. The EU introduces a new CO_2 emissions trading system in 2008 which drove up the prices of burning fossil fuels. Due to the closure of power plants at the beginning of the 2000s, there was not a need for capacity which also contributed to the steady price surge. In combination, there was a general increase in commodity prices in 2007 and 2008 which reached a peak in the middle of 2008. It was largely driven also by increased economic activity leading up to the 2008 financial crisis.

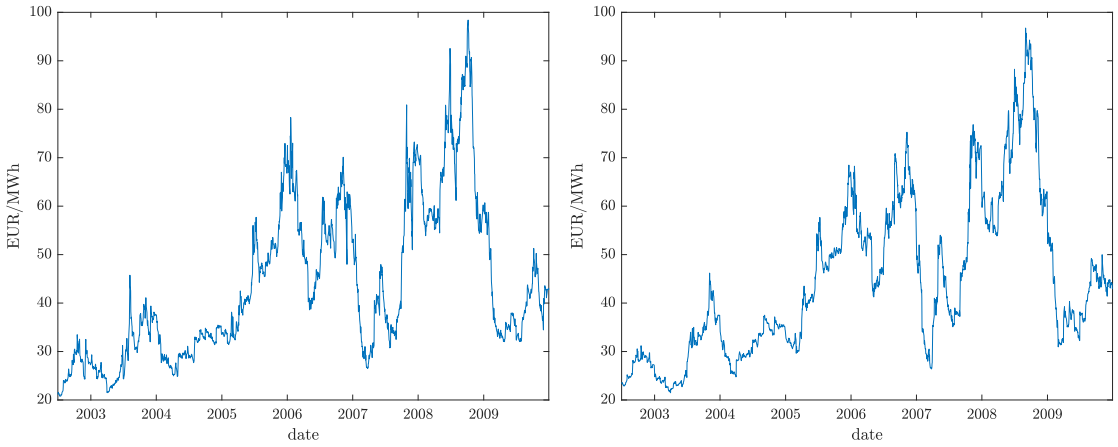
- **A sudden fall in price in late 2008**. In September 2008 the collapse of Lehman Brothers happened which led to a global recession that halted the great demand for energy commodities and prices fell. Additionally, the banking and debt crisis hit especially southern European countries hard hence participants were forced to find liquidity. This meant that CO_2 certificated were sold which drove down the price of CO_2 and subsequently power.

In the above, we have not even discussed the effect of meteorological variables and the gradual introduction of renewable energy sources and the effect on the spot prices.

E.1.2 Futures Prices

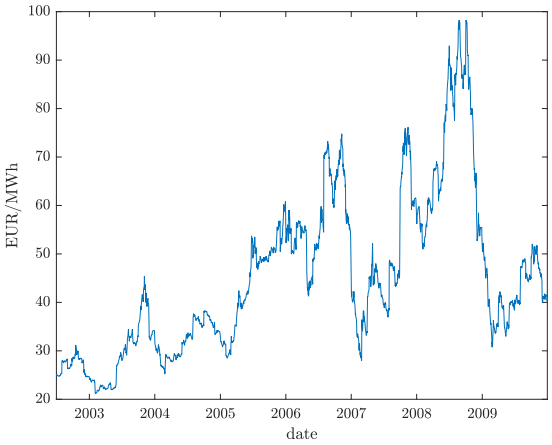
We will analyze EEX power futures closing prices on a monthly basis. These prices are available in the market for specific months, such as Jan-05 or Feb-05, but not all at once. Monthly contracts can only be traded for a limited period in the future. To limit the scope of our analysis and account for low liquidity for monthly contracts beyond four months, we have chosen to only consider contracts that have a maximum of four months between the current time t and the start of the delivery T_1 in a forward contract $F(t, T_1, T_2)$.
To construct price series that covers the entire period, we use contract rollover. This means that we consider the time to delivery instead of the specific time of delivery, e.g. instead of one price series for Jan-04 that is very short, we construct a series of contracts with one month to delivery, M1. We rollover at the end of the month such that when t is in December 2003, then the price signal from Jan-04 constitutes $F_{M1}(t)$. Likewise, when t is in Jan-04, then $F_{M1}(t)$ is the price series for Feb-04 in the period. In chapter F, we discuss the assumptions when we construct such a price series. The constructed series can be seen below and we note that many of the events described in section E.1.1, can also be seen directly and sometimes even amplified in the futures contracts. There is one major difference:

- **Strong seasonality**. For all of the futures, there is always an increase in the futures prices leading up to the winter. Indeed, this is because these contracts cover these, however, the realized spot prices in the same period do not seem to have a strong yearly seasonal signal though they should intuitively also follow the heating season. Here there is an obvious risk premium. Before the winter, there is a lot of uncertainty around how cold it will become and if the fuel storages are sufficient. Therefore, the market takes a premium for taking on the risk of the prices that might explode.

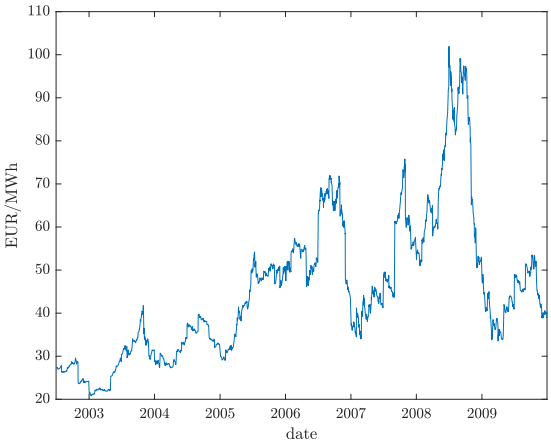


(a) M1 contract in the period considered.

(b) M2 contract in the period considered.



(c) M3 contract in the period considered.

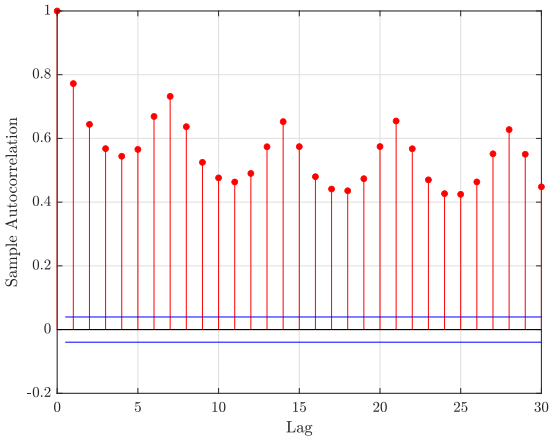


(d) M4 contract in the period considered.

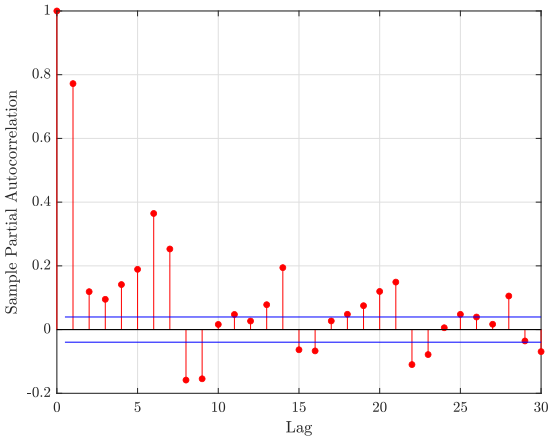
Figure E.2 – The constructed series of futures contracts based on contract rollover.

E.2 Explorative Data Analysis of the Spot Price Series

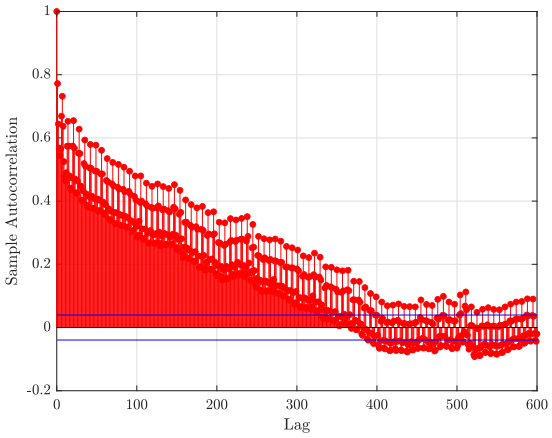
In the following, we will describe some characteristics of the spot price to motivate the models we consider.



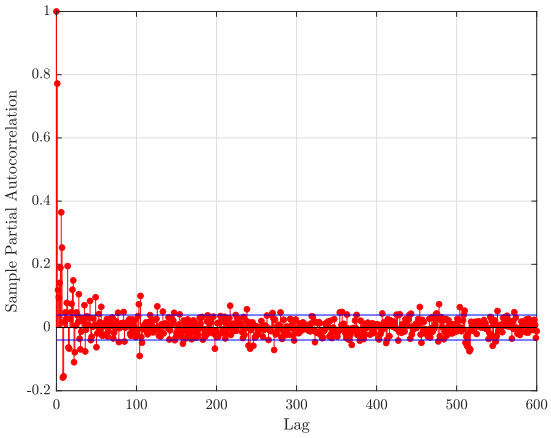
(a) Sample Autocorrelation Function.



(b) Sample Partial Autocorrelation Function.



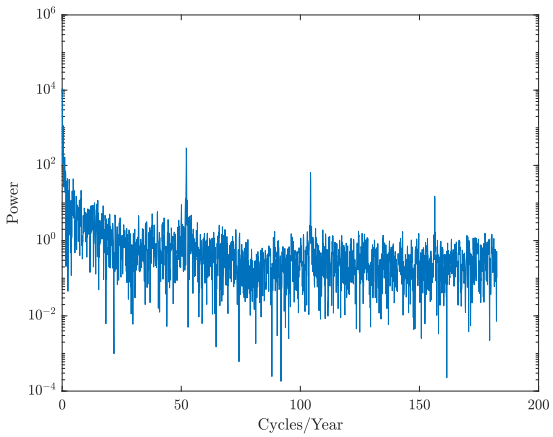
(c) Sample Autocorrelation Function.



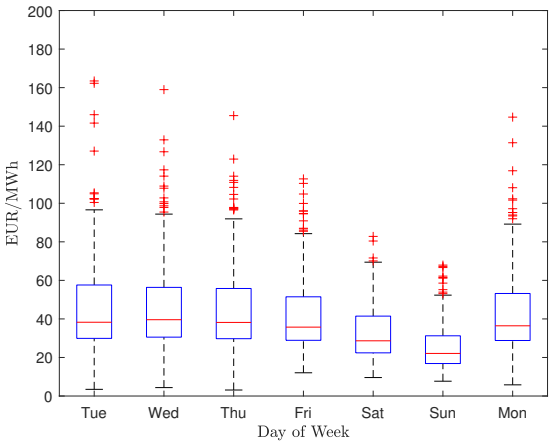
(d) Sample Partial Autocorrelation Function.

Figure E.3 – Analysis figures to find seasonality in the spot price series.

The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) plots in fig. E.3a and fig. E.3b reveal strong signals, with a distinct 7-day signal indicating a weekly seasonality in the data. As more lags are included in fig. E.3c and fig. E.3d, there is no apparent yearly seasonality observed. This is consistent with the findings of [9], which also analyzed this period for the German market and did not identify any yearly seasonality, despite its presence in other markets such as Nordpool. The strong weekly signal is further supported by the periodogram in fig. E.4a, which shows a peak at 52 cycles per year, corresponding to the number of weeks in a year, as well as its harmonics at 104 and 156. To further investigate this pattern, we analyzed the distribution of prices by individual days in fig. E.4b, and observed that prices are generally higher during weekdays and exhibit greater variation compared to weekends.



(a) Periodogram with cycles per year.



(b) Boxplot for the prices for each days of the week where we have limited to only look until a price of 200 EUR/MWh.

Figure E.4 – Analysis figures to find seasonality in the spot price series.

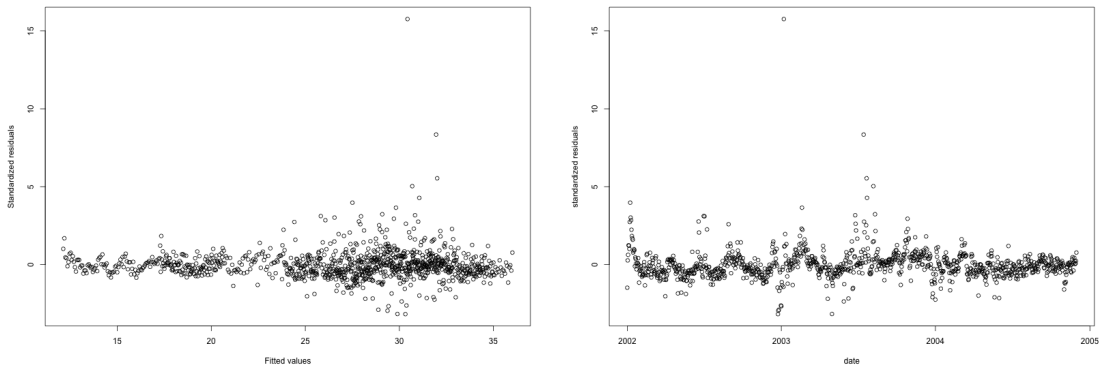
E.3 Analysis for the Short Period

In this section, we will construct a model for the short period. We will first find a model for the deterministic function, $\Lambda(t)$, and use it to identify the spikes. We will start with a

parsimonious model built directly on intuition from the statistical analysis. As written in section B.2.7, we will consider a model of the type:

$$Y_{t_i} = \beta t_i + \alpha_1 \mathbb{1}_{\{mon\}} + \alpha_2 \mathbb{1}_{\{tues\}} + \cdots + \alpha_7 \mathbb{1}_{\{sun\}} + \epsilon_{t_i}, \tag{E.3.1}$$

where $\epsilon_{t_i} \sim \mathcal{N}\left(0, \sigma^2\right)$. We fit the linear model using the built-in `lm` function in **R**. The parameters of the model can be found in table D.1 and a variety of residual plots are presented in fig. D.1. In this analysis, we focus on the standardized residuals plotted over time, as seen in fig. E.5b, and the fitted values, as seen in fig. E.5a. In fig. E.5b, we observe some spikes and strong mean reversion but no worrying signs. In fig. E.5a, we notice that the residual variance increases with the fitted values, suggesting the need for a transformation of Y_t as suggested in section 3.10 of [23]. Therefore, we introduce a log transformation.



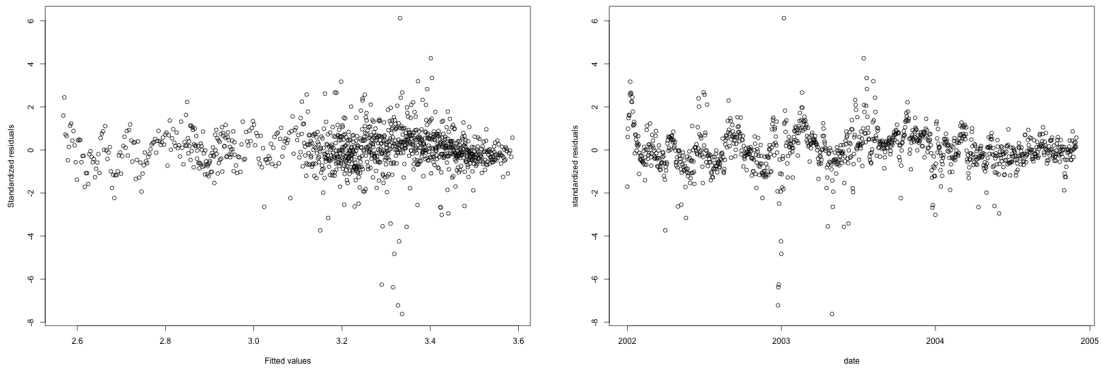
(a) Standardized residuals with fitted values. **(b)** Standardized residuals with time.

Figure E.5 – Selected figures to asses residuals of the model in eq. (E.3.1).

Consider the following log-transpose of the model:

$$\log\left(Y_{t_i}\right) = \beta t_i + \alpha_1 \mathbb{1}_{\{mon\}} + \alpha_2 \mathbb{1}_{\{tues\}} + \cdots + \alpha_7 \mathbb{1}_{\{sun\}} + \epsilon_{t_i}. \tag{E.3.2}$$

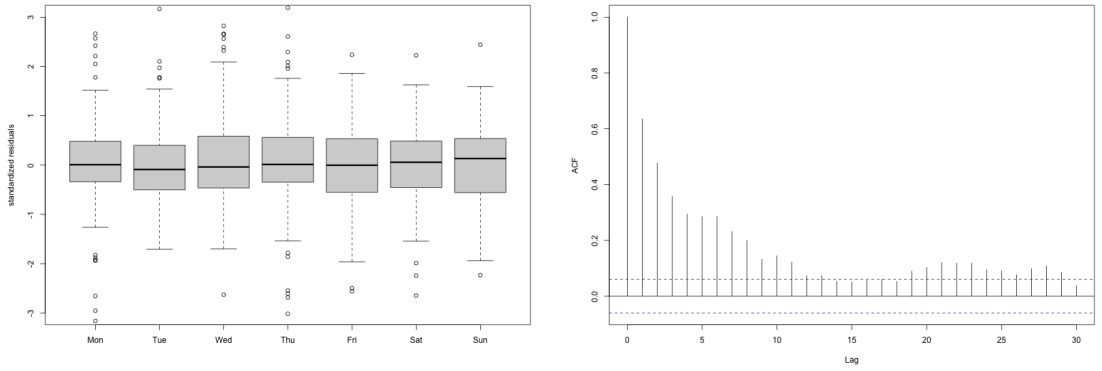
We can also write this as $Y_{t_i} = \exp\left(\beta t_i + \alpha_1 \mathbb{1}_{\{mon\}} + \alpha_2 \mathbb{1}_{\{tues\}} + \cdots + \alpha_7 \mathbb{1}_{\{sun\}}\right) \exp\left(\epsilon_{t_i}\right)$ and then we observe that out model is multiplicative. In this setting, we also see that the error, $\exp\left(\epsilon_{t_i}\right)$, becomes multiplicative and log-normal. The found parameters for the model in eq. (E.3.2) can be found in table D.3 with the full diagnostic plots can be seen in fig. D.2. We will only devote our attention to a few.



(a) Standardized residuals with fitted values. **(b)** Standardized residuals with time.

Figure E.6 – Selected figures to asses residuals of the model in eq. (E.3.1).

In fig. E.6a, we observe that the residuals do not seem to scale as much with the fitted values as before. However, in fig. E.6b, there is still some systematic variance and a spike signal that the model has not captured. Some of this systematic variance could be due to differences in trend or variance for each weekday. In appendix D.3, we attempted to expand the model to include a time trend for each weekday, however, this did not improve the residuals plot and the added terms were found to be insignificant. In fig. E.7a, we note that the interquartile range (IQR) and median differ slightly for each weekday, and we will consider a variance parameter for each weekday in the next subsection. Additionally, in fig. E.7b, we observe that there is still serial correlation present in the data that needs to be accounted for, which we will address in the next section.

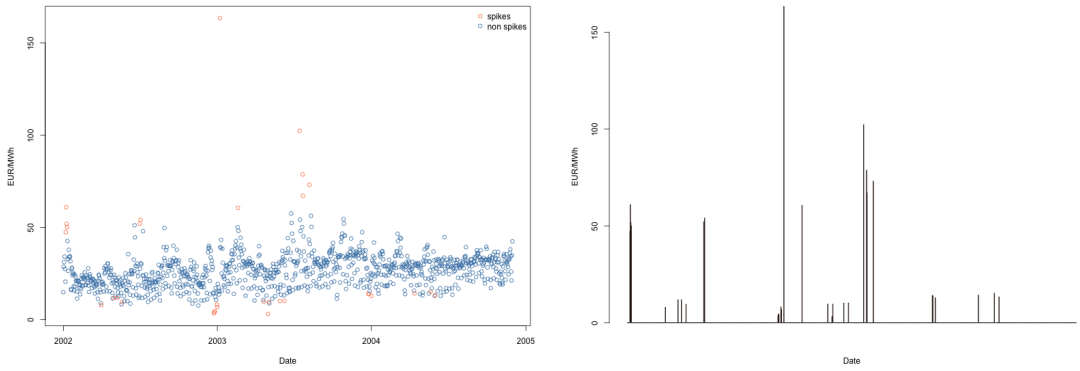


(a) Residuals with truncated y-limits. All spikes can be seen in fig. D.3 **(b)** The ACF of the standardized residuals.

Figure E.7 – Selected figures to asses residuals of the model in eq. (E.3.2)

E.3.1 Removal of the Spikes and Introduction of New Variance Structures

With the model in eq. (E.3.2), we now filter out the spikes. A simple threshold-based rule is applied, stipulating that a point is considered a spike if the standardized residual is smaller than -2.5 or larger than 2.5. In fig. E.8a, the located spikes using this approach are plotted with the rest of the series, and in fig. E.9b, the prices at the located spikes are plotted as bars.



(a) The located spikes in the data.

(b) Barplot of the price value of the located spikes.

Figure E.8 – Selected figures to asses residuals of the model in eq. (E.3.1).

We can now **remove** the data points with the identified spikes as we will use them to find the parameters for the spike signal. We will now **refit** our model on the dataset where the spikes are removed. We then obtain the following parameter estimates in table E.1.

	2.5 %	β	97.5 %
β	0.000330	0.000372	0.000415
α_1	3.101	3.143	3.185
α_2	3.141	3.182	3.224
α_3	3.146	3.188	3.230
α_4	3.159	3.201	3.243
α_5	3.066	3.107	3.148
α_6	2.806	2.847	2.888
α_7	2.523	2.563	2.604
σ^2	0.0421	0.0458	0.0501

Table E.1 – Parameter estimates for eq. (E.3.2) fitted on the data where the identified spikes in fig. E.8a have been removed

A Variance Parameter for Each Weekday

In fig. E.7a, we also saw that the distribution of the weekdays seemed to differ slightly. Therefore, we will expand the model structure and have a separate variance parameter for each weekday as also considered in section 4.1.3 in [24]:

$$\log\left(Y_{t_i,j}\right) = \beta t_i + \alpha_1 \mathbb{1}_{\{mon\}} + \alpha_2 \mathbb{1}_{\{tues\}} + \cdots + \alpha_7 \mathbb{1}_{\{sun\}} + \epsilon_{t_i,j}, \tag{E.3.3}$$

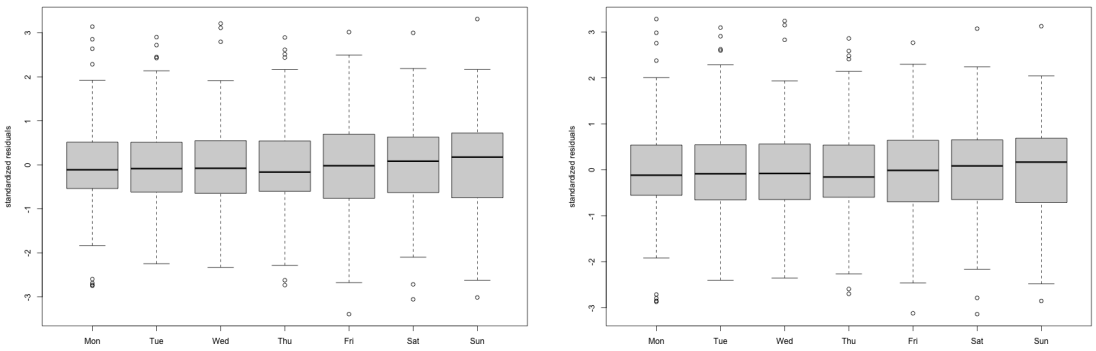
where

$$\epsilon_{t_i,j} \sim \mathcal{N}\left(0, \sigma_j^2\right), \qquad j \in \{mon, tue, \ldots, sun\}. \tag{E.3.4}$$

Here we use the generalized least squares function, `gls`, from the `nmle` package and use the argument `weights = varIdent(form = 1|factor(weekday))`. As the models in eq. (E.3.2) and eq. (D.2) are nested, we can use the log-likelihood ratio test for the null-hypothesis:

$$H_0 : \sigma^2_{mon} = \sigma^2_{tues} = \cdots = \sigma^2_{sun} \tag{E.3.5}$$

Based on the log-likelihood ratio test, we accept null hypothesis as it can be seen in table D.6 in the appendix.
Graphically, we can also see that the additional variance parameters do not seem to help much when we compare the normalized residuals in fig. E.9a from the model with simple variance structure to the residuals for the model with a variance parameter for each week, fig. E.9b.



(a) Residuals for each weekday with removed spikes and model eq. (E.3.2).

(b) Residuals for each weekday with removed spikes and the model eq. (D.2) with a variance parameter for each weekday.

Figure E.9 – Selected figures to asses residuals of the model in eq. (E.3.1).

Account for the Autocorrelation in the Residuals

To account for the strong ACF signal in fig. E.7b, we introduce an AR(1) correlation structure for the residuals such that our model is now as in eq. (E.3.2), but we introduce a ρ such that:

$$\epsilon_i \sim \mathcal{N}(0, \Lambda), \qquad \Lambda = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}. \tag{E.3.6}$$

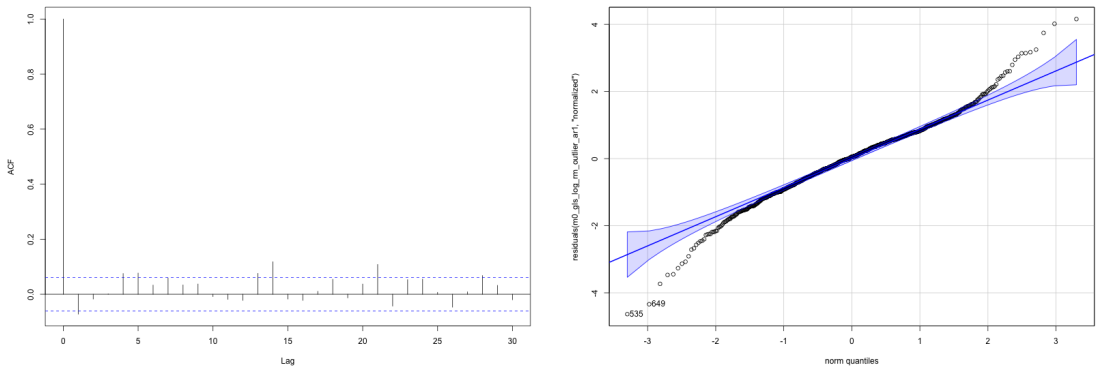
This way of factorizing the covariance matrix into a variance parameter and correlation parameter is common in the *gls* framework [25]. We specify this using the argument `correlation = corAR1(form = t))`. When we fit this model, we obtain the parameters specified in table E.2.

	2.5 %	β	97.5 %
β	0.000274	0.000377	0.000479
α_1	3.065	3.132	3.199
α_2	3.114	3.181	3.248
α_3	3.122	3.189	3.256
α_4	3.122	3.189	3.256
α_5	3.036	3.103	3.170
α_6	2.774	2.841	2.907
α_7	2.493	2.560	2.626
σ^2	0.0424	0.0492	0.0572
ρ	0.651	0.699	0.742

Table E.2 – Parameter estimates for eq. (E.3.2) fitted on the data where the spikes have been removed and AR(1) structure in residuals is imposed as in eq. (E.3.6).

Note that in the following, we will use the *normalized* residuals as defined in [25], section 5.3.4, $\boldsymbol{r}_i = \widehat{\sigma}^{-1}\left(\widehat{\boldsymbol{\Lambda}}_i^{-1/2}\right)^T(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)$. These should follow a standard normal if the model is correctly specified. In fig. E.10a, we see that the correlation parameter has removed the signal in the ACF. Additionally, when we make an F-test for the auto-correlation parameter, for which we see that the model is significant as it can be seen in appendix D.5.

The model in fig. E.10b is not perfect, as evidenced by the residuals. However, given the limitations of the available data, we accept that there is still some amount of variance that cannot be fully described by the model. Additional diagnostic plots can be found in fig. D.5. The residuals in the original domain, $\boldsymbol{y}_i - \exp\widehat{\boldsymbol{y}}_i$, are calculated. On days where spikes were identified, we use the residuals from the previous day.



(a) ACF with normalized residuals.

(b) Qqplot for the normalized residuals

Figure E.10 – Selected residuals figures for the model eq. (E.3.2) with variance structure eq. (E.3.6)

E.3.2 Parameter Estimation for the Spike process

To estimate the parameters of the spike process, we took the pragmatic approach as is also used in [2], we will discuss alternatives in the discussion chapter F. The idea is to define proper jumps. In fig. E.9b we have the price value of the spikes, $\hat{s}_t^{J,\text{raw}}$. With our

found model fitted without the spikes, we can find fitted values and transform them to the expected prices, $\hat{\mu}_t$. We now introduce ffl as the jump with sign. We can now introduce proper jumps as in notion 3.5.2, [2]:

Notion 3.5.2. Proper jumps. We will call a positive jump in $\hat{S}_t^J$ proper if $\hat{S}_t^J \neq 0$ and if $\hat{S}_t^J \geq \hat{S}_{t-1}^J$. We will call a negative jump proper if $\hat{S}_t^J \neq 0$ and if $\hat{S}_t^J \leq \hat{S}_{t-1}^J$.

The frequency of the proper jumps is an estimate of λ; p and q can be calculated as the proportions of jumps with positive and negative spikes respectively; the expectation of an exponential density $\eta_1 e^{-\eta_1 x}$ is $1/\eta_1$ hence using the mean of the positive proper jumps, we can find η_1 and likewise for η_2 for the negative jumps. To estimate the spike mean reversion parameter β, we use **algorithm 3.5.3** in [2] with $k_\beta = 0.20$. When we do that, we find the following parameters.

λ	p	q	η_1	η_2	β
40.962	0.3462	0.6538	0.0200	0.0558	1.0901

Table E.3 – Parameter estimates for the spike process

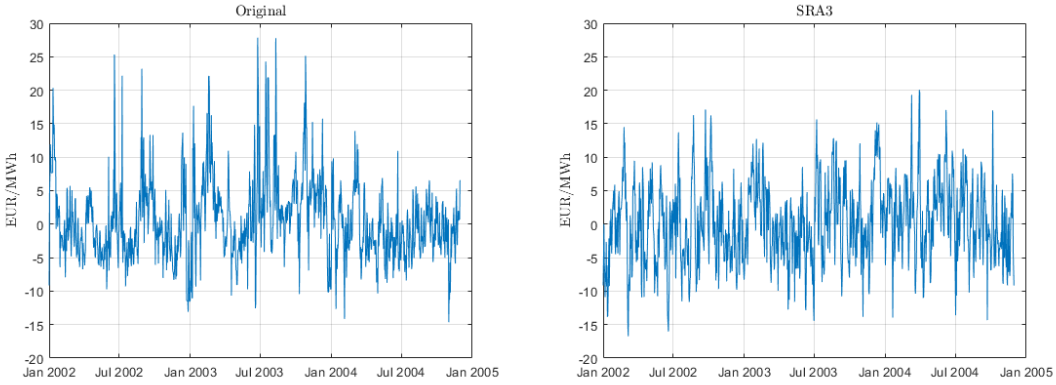
E.3.3 Parameter Estimation for the Ornstein–Uhlenbeck process

We see in table E.4 that the SRA3 method is very close to the exact MLE method. The Euler-Maruyama method on the other hand is over 15% away from the exact MLE estimates which indicates the data is not sampled frequently enough for the Euler-Maruyama method to work.

	Exact MLE	SRA3 MLE	EM MLE
μ	0.3606	0.3597	0.3028
σ	5.1208	5.2284	4.3223

Table E.4 – Parameter estimates for the Ornstein-Uhlenbeck process

As seen in fig. E.11a, the data used to fit the Ornstein Uhlenbeck process is displayed. In fig. E.11b, we can observe the reconstructed signal which utilizes the parameter estimates from the SRA3 method. The original signal exhibits many small fluctuations between -5 and 5, as well as larger fluctuations with a greater amplitude. However, the reconstructed signal does not possess as many large fluctuations as the original and appears to have a larger average fluctuation. This could suggest that the original signal may be better modeled by utilizing two Ornstein Uhlenbeck processes. Despite this, we will accept the reconstructed signal.



(a) Original signal.

(b) Recreated signal using the SRA3 parameters.

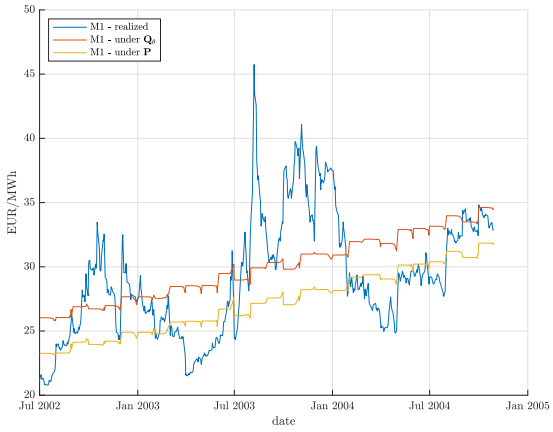
E.3.4 Estimating the Risk Premium

We will now calculate the risk premium that we presented in section B.2.5 and, specifically, we will find the constant risk premium, see eq. (B.2.33), for each future of the granularities; M1, M2, M3, and M4. We find the measure change parameters θ using the Sequential Least SQuares Programming (SLSQP) Algorithm in the `scipy.minimize` optimization packages in `Python`. When we do that, we obtain the following parameters estimates:

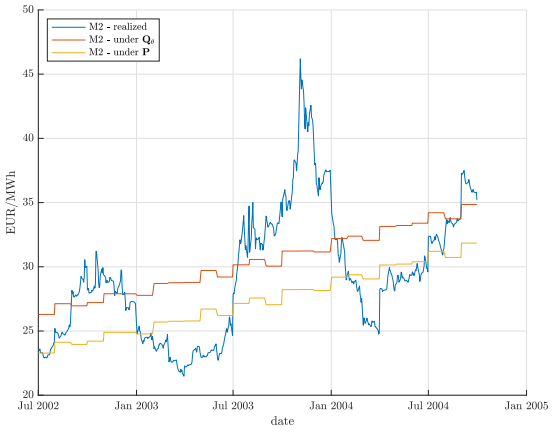
	M1	M2	M3	M4
θ_B	1.097	1.183	1.064	0.7659
θ_L	0.5716	0.6173	0.5695	0.4373

Table E.5 – Parameter estimates constant risk premia.

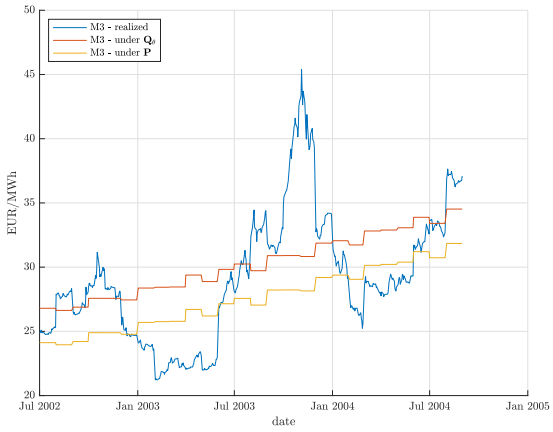
To understand the meaning of these parameters, consider first section E.3.4 where we see the price of actual daily closing prices of the future along with the price under the model. We first notice that the prices under $\mathbb{P}$ are always lower than the prices under $\mathbb{Q}_\theta$ which is because the estimated M1 risk premium is positive. The market premium is also positive for M2, M3, and M4, however, the value of the market premium differs slightly for each month. Graphically, we can see that as an increase/decrease in the distance between the price under $\mathbb{P}$ and under $\mathbb{Q}_\theta$.
The graphical representation looks peculiar in that it quite frankly looks like the model fits the futures prices poorly. The obvious seasonality in the futures data is not anticipated by our model. The main reason for this is due to the assumption of a constant risk premium over the year which seems inappropriate. We will discuss this more in chapter F.
In table E.5, we see that θ_B and θ_L seem to increase slightly from M1 to M2 and then fall from M3 to M4 which is in alignment with the results in [4]. From a market perspective, the M1 premium should also be lower than M2 as all participants will have more information and thus the risk associated with the contract is lower. The fall in risk premium from M2 to M4, is analyzed in [5] and the find that the phenomenon is described by the risk aversion by buyers and producers of electricity. More buyers want to hedge the price risk on the M2 forward horizon and are whiling to pay the additional market risk premium to make that happen. On a longer horizon, there will be fewer buyers and hence the risk premium falls.



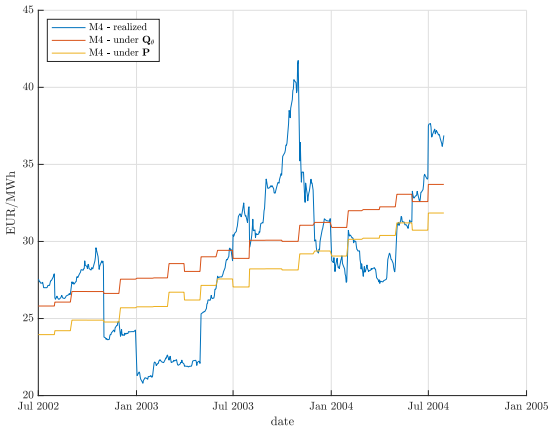
(a) M1 contract with the risk premium.



(b) M2 contracts with the risk premium.



(c) M3 contracts with the risk premium.



(d) M4 contracts with the risk premium.

Figure E.12 – The futures contract prices along with the price of the futures under the equivalent measure $\mathbb{Q}_\theta$ and direct pricing measure $\mathbb{P}$.

F | Discussion

F.0.1 The Construction of the Futures Price Series

Construction of futures prices series with rollover. When we use contract rollover and use M_1 instead of specific contracts as Jan-04, then we can construct longer series but this comes with some trade-offs. First of all the dynamics of winter M_1 contracts could be very different than summer M_1 contracts. This could also mean there could be drastic shifts in the price series when we roll over to a new contract. This is very evident for M_3 and M_4 contracts that almost go vertical at some shifts. Additionally, as noted in [26] for a specific contract the time to delivery varies from $T_1 - t$ being almost a month to having delivery the next day. This means much for the information available and how much of it is priced in. On top of that, most participants would close their positions some days prior to the end of the trading period. Therefore, there might be very little activity in the end. To mitigate this, one could instead define the contract rollover on the tick volume such that the M_1 is rolled over when there is more activity on the next monthly contract.

F.0.2 The Estimation of the Spike Parameters

We estimate the spike parameters in a pragmatic approach as in [2], however, there exist alternative methods for the estimation of Kou model as presented in [27]. They present methods using Monte Carlo, generalized methods of moments, and cumulant matching. In section 6.5.1 in [26], they show how extreme value theory can also be used to disentangle the spikes from the signal and then subsequently estimate parameters.

F.0.3 The Simplicity of the Model for $\Lambda(t)$

The model for the deterministic component, $\Lambda(t)$, is very simple only with intercepts for each weekday, a linear component in time, and the variance structure of eq. (E.3.6). The reason for this simplicity is for the model to represent the intended i.e. the obvious knowledge available to everyone. Additionally, as the model had to work for M1, M2, M3, and M4, we were also restricted to a simpler class of models.

F.0.4 Market Premium as a Yearly Function and Information Premium

In section E.3.4, we could see that the assumption of constant risk premium seemed quite poor. One could consider a time-dependent function as $\theta_B(t) = \alpha + \beta \cos \frac{2\pi t}{365}$ and likewise for θ_L which would be a recommendation for further studies. In [2] [26] [4] they consider how one can formalize the additional information that idea that information about the future should be priced in. In appendix C, we show how that could be formalized in case one used the model we consider in this report. It could be interesting to continue this in future work.

G | Conclusion

In this study, we analyzed methods to quantify the risk premium in electricity futures contracts in the German power market. To find the market risk premium, we used a change of measure from the direct pricing measure, $\mathbb{P}$, to an equivalent measure, $\mathbb{Q}_\theta$, parameterized by a vector of risk premium parameters, θ. We did this using the methods introduced in [2] where one needs to find a model for the spot prices, S_t which has some constituent parts, $S_t = \Lambda(t) + X_t + Y_t$. Unlike [2], we find that for the deterministic part $\Lambda(t)$, there is no yearly nor semi-yearly component which is also in alignment with other authors [9] that studied the same period. To describe the diffusive component, X_t, we used an Ornstein-Uhlenbeck process and showed how the recent SR3 method is more stable and useful to estimate parameters than the standard EM method. This meant that we could avoid applying simulated maximum likelihood which is much more computationally demanding. We used directly the methods from [2] to find parameters for the spike component, Y_t, and then were able to determine θ. In the period from 2002 to 2004, we find positive risk premium parameters for all M1, M2, M3, and M3 futures contracts in the German Power market. In accordance with [4], we find a larger risk premium for the M2 contract, however, we also saw that one should consider a yearly seasonality for the premium as the assumption of a constant risk premium proved too poor. For further studies, it would also be interesting to dive further into the concept of the *information premium* that we laid the groundwork for in appendix C.

Bibliography

[1] (2019) The liberalisation of electricity markets in germany history, development and current status. [Online]. Available: https://www.agora-energiewende.de/en/publications/the-liberalisation-of-electricity-markets-in-germany/

[2] R. Biegler-König, "Information premium on electricity markets: A new spot-forward relationship for non-storable underlyings," Ph.D. dissertation, May 2013. [Online]. Available: https://duepublico2.uni-due.de/receive/duepublico_mods_00031052

[3] F. E. Benth, J. S. Benth, and S. Koekebakker, *Stochastic modelling of electricity and related markets*. World Scientific, 2008, vol. 11.

[4] F. E. Benth, R. Biegler-König, and R. Kiesel, "An empirical study of the information premium on electricity markets," *Energy Economics*, vol. 36, pp. 55–77, 2013.

[5] F. E. Benth, Á. Cartea, and R. Kiesel, "Pricing forward contracts in power markets by the certainty equivalence principle: Explaining the sign of the market risk premium," *Journal of banking & finance*, vol. 32, no. 10, pp. 2006–2021, 2008.

[6] S. G. Kou, "A jump-diffusion model for option pricing," *Management science*, vol. 48, no. 8, pp. 1086–1101, 2002.

[7] D. Applebaum, *Lévy processes and stochastic calculus*. Cambridge university press, 2009.

[8] B. Øksendal and A. Sulem, *Stochastic Control of jump diffusions*. Springer, 2005.

[9] W. Rafal, "Modelling and forecasting electricity loads and prices," 2006.

[10] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, ser. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2019.

[11] C. Rackauckas, "An intuitive introduction for understanding and solving stochastic differential equations," 2017.

[12] D. Kressner, *Advanced Numerical Analysis*. EPFL, 2015.

[13] A. Rößler, "Rooted tree analysis for order conditions of stochastic runge-kutta methods for the weak approximation of stochastic differential equations," *Stochastic Analysis and Applications*, vol. 24, no. 1, pp. 97–134, 2006.

[14] ——, "Runge–kutta methods for the strong approximation of solutions of stochastic differential equations," *SIAM Journal on Numerical Analysis*, vol. 48, no. 3, pp. 922–952, 2010.

[15] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, ser. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2011. [Online]. Available: https://books.google.dk/books?id=BCvtssom1CMC

[16] D. B. Hernandez and R. Spigler , "A-stability of runge-kutta methods for systems with additive noise," *BIT Numerical Mathematics*, vol. 32, 1992.

[17] S. S. Artemiev and T. A. Averina, *Numerical Analysis of Systems of Ordinary and Stochastic Differential Equations*. Berlin, New York: De Gruyter, 1997. [Online]. Available: https://doi.org/10.1515/9783110944662

[18] E. Buckwar, M. Riedler, and P. Kloeden, "The numerical stability of stochastic ordinary differential equations with additive noise," *Stochastics and Dynamics*, vol. 11, 11 2011.

[19] J. K. Møller and H. Madsen, "From state dependent diffusion to constant diffusion in stochastic differential equations by the lamperti transform," 2010.

[20] C. Rackauckas and Q. Nie, "Stability-optimized high order methods and stiffness detection for pathwise stiff stochastic differential equations," pp. 1–8, 2020.

[21] S. M. Iacus, *Simulation and inference for stochastic differential equations: with R examples*. Springer, 2008, vol. 486.

[22] A. R. Pedersen, "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations," *Scandinavian journal of statistics*, pp. 55–71, 1995.

[23] H. Madsen and P. Thyregod, *Introduction to general and generalized linear models*. CRC Press, 2010.

[24] A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith *et al.*, *Mixed effects models and extensions in ecology with R*. Springer, 2009, vol. 574.

[25] J. Pinheiro and D. Bates, *Mixed-effects models in S and S-PLUS*. Springer science & business media, 2006.

[26] F. E. Benth, V. A. Kholodnyi, and P. Laurence, "Quantitative energy finance," *Modelling, pricing, and hedging in energy and commodity markets. Springer*, 2014.

[27] A. Karimov, "Numerical implementation and parameter estimation under kou model," in *Identifying Stock Market Bubbles*. Springer, 2017, pp. 73–96.

[28] K. Bandara, R. J. Hyndman, and C. Bergmeir, "Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns," 2021. [Online]. Available: https://arxiv.org/abs/2107.13462

[29] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition," *J. Off. Stat*, vol. 6, no. 1, pp. 3–73, 1990.

Appendices

A Appendix for derivations linked to chapter B

A.1 Forward Curve for a specific date

In the following, we will cover some fundamental properties of a standard forward curve. Consider a generic forward figure of the forward curves prices quoted on 01-10-2021, dd-mm-yyyy:

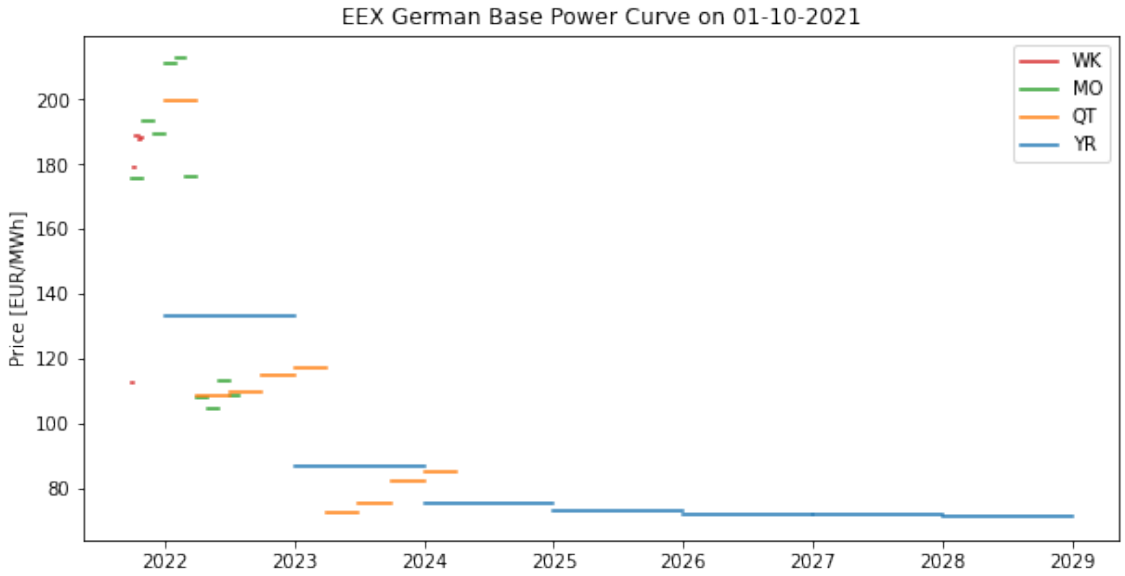


Figure A.1 – Closing price of EEX base load power futures quoted on 01-10-2021

We see that the contracts with shorter maturities seems to be increase for the first two years into the future which is a sign of *contango*. On the the other hand, there seems to be a general downward trend for contracts with longer time to maturities which means they are in *backwardation*.

A.2 Derivation of Expression for the Log-Moments

$$
\begin{aligned}
\psi_{L_{\mathrm{d}s}}(u) &= \log\left(\exp\left(\mathrm{d}s\lambda\int_{\mathbb{R}}\left(e^{uy}-1\right)F(dy)\right)\right) \\
&= \mathrm{d}s\lambda\int_{\mathbb{R}}\left(e^{uy}-1\right)\left(p\eta_1 e^{-\eta_1 y}\mathbb{1}_{\{y\geq 0\}}+q\eta_2 e^{\eta_2 y}\mathbb{1}_{\{y<0\}}\right)dy \\
&= \mathrm{d}s\lambda\eta_1 p\int_0^\infty\left(e^{uy}-1\right)e^{-\eta_1 y}dy+\mathrm{d}s\lambda\eta_2 q\int_{-\infty}^0\left(e^{uy}-1\right)e^{\eta_2 y}dy \qquad \text{(A.1)} \\
&= \mathrm{d}s\lambda\eta_1 p\int_0^\infty\left(e^{(u-\eta_1)y}-e^{-\eta_1 y}\right)dy+\mathrm{d}s\lambda\eta_2 q\int_{-\infty}^0\left(e^{(u+\eta_2)y}-e^{\eta_2 y}\right)dy \\
&= \frac{-\mathrm{d}s\lambda pu}{u-\eta_1}+\frac{\mathrm{d}s\lambda qu}{u+\eta_2}
\end{aligned}
$$

B Appendix for derivations linked to chapter C

B.1 SRK of strong order 1.0 with scalar noise

The stochastic Runge-Kutta methods with a strong order of 1.0 and scalar noise has the following form if the Brownian motion only has the dimension of 1.

$$\begin{aligned} x(t_{k+1}) = x(t_k) &+ \sum_{i=1}^{s} \alpha_i \mathbf{f}\left(t_k + c_i^{(0)}\Delta t, H_i^{(0)}\right)\Delta t \\ &+ \sum_{i=1}^{s} \left(\beta_i^{(1)} I_{(1)} + \beta_i^{(2)}\sqrt{\Delta t}\right) \mathbf{g}\left(t_k + c_i^{(1)}\Delta t, H_i^{(1)}\right) \end{aligned} \tag{B.1}$$

$$\begin{aligned} H_i^{(0)} &= x(t_k) + \sum_{j=1}^{s} A_{ij}^{(0)} \mathbf{f}\left(t_k + c_j^{(0)}\Delta t, H_j^{(0)}\right)\Delta t + \sum_{j=1}^{o} B_{ij}^{(0)} \mathbf{g}\left(t_k + c_j^{(1)}\Delta t, H_j^{(1)}\right) I_{(1)}, \\ H_i^{(1)} &= x(t_k) + \sum_{j=1}^{s} A_{ij}^{(1)} \mathbf{f}\left(t_k + c_j^{(0)}\Delta t, H_j^{(0)}\right)\Delta t + \sum_{j=1}^{s} B_{ij}^{(1)} \mathbf{g}\left(t_k + c_j^{(1)}\Delta t, H_j^{(1)}\right) \frac{I_{(1,1)}}{\sqrt{\Delta t}} \end{aligned} \tag{B.2}$$

The integrals $I_{(1)}$ and $I_{(1,1)}$ are given below.

$$I_{(1)} = \mathrm{dB}(t) \tag{B.3}$$

$$I_{(1,1)} = \frac{1}{2}(I_{(1)}^2 - \Delta t) \tag{B.4}$$

Butcher tableau can be written on the following form.

$$\begin{array}{c|c|cc} c^{(0)} & A^{(0)} & B^{(0)} & \\ \hline c^{(1)} & A^{(1)} & B^{(1)} & \\ \hline & \alpha^T & \beta^{(1)^T} & \beta^{(2)^T} \end{array} \tag{B.5}$$

We have implemented a method with strong order 1 and weak order 2 for which the butcher tableau is given in B.1.

$$\begin{array}{c|ccc|ccc|ccc} 0 & & & & & & & & & \\ 1 & 1 & & & 0 & & & & & \\ 0 & 0 & 0 & & 0 & 0 & & & & \\ \hline 0 & & & & & & & & & \\ 1 & 1 & & & 1 & & & & & \\ 1 & 1 & 0 & & -1 & 0 & & & & \\ \hline & \frac{1}{2} & \frac{1}{2} & 0 & 1 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{array}$$

Figure B.1 – Caption

C Information Premium

As stipulated in section B.1.4, one could also consider a filtration, $\mathcal{G}_t$, that includes some information about the possible level, K, of the electricity in addition to the historical information in $\mathcal{F}_t$, e.g. $\mathcal{G}_t \subseteq \mathcal{F}_t \cup \sigma\left(\mathbb{1}_{S_{T_\Gamma} \geq K}\right)$. If this piece of information is relevant at some timeT_Γ, then we can define the information premium,

$$I^{\mathbb{Q}}_{\mathcal{G}_t}(t, T, T_\Gamma) = F^{\mathbb{Q}}_{\mathcal{G}_t}(t, T) - F^{\mathbb{Q}}_{\mathcal{F}_t}(t, T), \tag{C.1}$$

which we in practice will understand as the difference between the forward prices under the market filtration $\mathcal{G}_t$ and historical filtration $\mathcal{H}_t$. Explicitly, we will assume that the observed forward prices are calculated by the market participants and correspond to $\mathbb{E}^{\mathbb{Q}}\left[S_T \middle| \mathcal{G}_t\right]$. Indeed, the expected information premium under $\mathcal{F}_t$ is none:

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}\left[I_{\mathcal{G}_t} \middle| \mathcal{F}_t\right] &= \mathbb{E}^{\mathbb{Q}}\left[F^{\mathbb{Q}}_{\mathcal{G}_t}(t, T) - F^{\mathbb{Q}}_{\mathcal{F}_t}(t, T) \middle| \mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[S_T \middle| \mathcal{G}_t\right](t, T) \middle| \mathcal{F}_t\right] \\ &= 0. \end{aligned} \tag{C.2}$$

where we have used the definition $F^{\mathbb{Q}}_{\mathcal{G}_t} = \mathbb{E}^{\mathbb{Q}}\left[S_T \middle| \mathcal{G}_t\right]$ and the tower property.

All this formalism is to ensure that we can correctly partition the difference from forecasted spot prices to futures seen in the market. The relation between the information premium and risk premium can be summarized.

C.1 Information Premium and Risk Premium

With the historical filtration of spot prices $\mathcal{F}_t$, the filtration with additional information $\mathcal{G}_t$, the price measure $\mathbb{Q}$ with market risk, and measure without thoughts on risk compensation $\mathbb{P}$, we introduce the risk premium

$$R^{\mathbb{Q}}_{\mathcal{G}_t} = R^{\mathbb{Q}}_{\mathcal{F}_t} + I^{\mathbb{Q}}_{\mathcal{G}_t} - I^{\mathbb{P}}_{\mathcal{G}_t}. \tag{C.3}$$

To understand this, consider a planned outage in the future. In $R^{\mathbb{Q}}_{\mathcal{F}_t}$ this additional risk not price and $I^{\mathbb{P}}_{\mathcal{G}_t}$ would give the additional spot price increment that a historically planned outage has given rise to. $I^{\mathbb{Q}}_{\mathcal{G}_t}$ would include the anticipated increment to the spot and how much the market participants should be compensated for this additional risk of not knowing exactly how much more the price might increase.

C.2 Information Premium under our Spot Price Model

In the following, let $\mathcal{G}_t$ be the market filtration with additional information $\mathcal{G}_t \subseteq \mathcal{H}_t = \mathcal{F}_t \cup \sigma\left(L_{T_\Gamma}\right)$ where L_{T_Γ} is Lévy information at time $T_\Gamma > T$. Explicitly, as $\sigma\left(L_{T_\Gamma}\right)$ is only Lévy information, it will essentially only affect the underlying driving noise L_{T_Γ}. Specifically for $L_{T_\Gamma} \perp X_t$ and $\mathbb{E}\left[X_t | \sigma\left(T_\Gamma\right)\right] = \mathbb{E}\left[X_t\right]$ as X_t is not $\sigma\left(T_\Gamma\right)$-measurable, then:

$$\mathbb{E}\left[X_t | \mathcal{G}_t\right] = \mathbb{E}\left[X_t | \mathcal{F}_t \cup \sigma\left(L_{T_\Gamma}\right)\right] = \mathbb{E}\left[X_t | \mathcal{F}_t\right] \tag{C.4}$$

This is a useful observation when we consider the information premium. Consider for now some $T \leq T_1$, then we write the information premium as:

$$I_{\mathcal{G}}^{\mathbb{Q}}(t,T;T_\Upsilon) = F_{\mathcal{G}_t}^{\mathbb{Q}}(t,T) - F_{\mathcal{F}_t}^{\mathbb{Q}}(t,T) \tag{C.5}$$

Given the considerations above and the fact that $\Lambda(t)$ is deterministic, we see that the additional information under $\mathcal{G}_t$, will only be relevant for the expectation of the Lévy increments in B.2.3, hence, for now, we can write:

$$\begin{aligned} I_{\mathcal{G}}^{\mathbb{Q}}(t,T;T_\Upsilon) =& \mathbb{E}\left[Y_t \mathrm{e}^{-\beta(T-t)} + \int_t^T \mathrm{e}^{-\beta(T-s)} dL_s \mid \mathcal{G}_t\right] \\ &- \mathbb{E}\left[Y_t \mathrm{e}^{-\beta(T-t)} + \int_t^T \mathrm{e}^{-\beta(T-s)} dL(s) \mid \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)} dL_s \mid \mathcal{G}_t\right] - \mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)} dL(s) \mid \mathcal{F}_t\right] \end{aligned} \tag{C.6}$$

We now have to calculate the expectations of the above. Consider, therefore, only $\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)} \mathrm{d}L(s) \mid \mathcal{F}_t\right]$ and use Fubini's theorem to switch integration and expectation:

$$\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)} dL(s) \middle| \mathcal{F}_t\right] = \int_t^T \mathrm{e}^{-\beta(T-s)} \mathrm{d}\mathbb{E}\left[L(\mathrm{d}s) \mid \mathcal{F}_t\right] \tag{C.7}$$

To calculate $\mathbb{E}\left[L(\mathrm{d}s) \mid \mathcal{F}_t\right]$, we need the log-moment generating function. In the following we write $\mathbb{E}\left[L(\mathrm{d}s)\right]$ to simplify notation. The log-moment generating function can be written as:

$$\psi_{L_{\mathrm{d}s}}(u) = \log \mathbb{E}\left[e^{uL_{\mathrm{d}s}}\right] \tag{C.8}$$

where we have the relationship:

$$\left.\frac{\mathrm{d}\psi_{L_{\mathrm{d}s}}(u)}{\mathrm{d}u}\right|_{u=0} = \mathbb{E}\left[L_{\mathrm{d}s}\right] \tag{C.9}$$

It can be shown that (sec. 1.3 [7]), for a compound Poisson process, the moment generating function is:

$$\mathbb{E}\left[e^{uL_t}\right] = \exp\left(t\lambda \int_{\mathbb{R}} \left(e^{uy} - 1\right) F(\mathrm{d}y)\right) \tag{C.10}$$

where $F(x)$ is the distribution of the jumps which in our case is $f_D(x)$. Indeed, as we know $\psi_{L_{\mathrm{d}s}}(u) = \log \mathbb{E}\left[e^{uL_t}\right]$, we can now derive $\psi_{L_{\mathrm{d}s}}(u)$ using the expression above. In appendix A.2, we show the derivations, and the ultimate result is that:

$$\psi_{L_{\mathrm{d}s}}(u) = \lambda\left(\frac{qu}{u+\eta_2} - \frac{pu}{u-\eta_1}\right)\mathrm{d}t \tag{C.11}$$

$$\psi'_{L_1}(u) = \left[\frac{-\lambda p}{u-\eta_1} + \frac{\lambda p u}{(u-\eta_1)^2} + \frac{\lambda q}{u-\eta_2} + \frac{-\lambda q u}{(u-\eta_2)^2}\right]\mathrm{d}t \tag{C.12}$$

with

$$\left.\frac{\mathrm{d}\psi_{L_{\mathrm{d}s}}(u)}{\mathrm{d}u}\right|_{u=0} = \lambda\left[\frac{p}{\eta_1} - \frac{q}{\eta_2}\right]\mathrm{d}t \tag{C.13}$$

hence we now have the result:

$$\mathbb{E}\left[L_{\mathrm{d}s}\right] = \left.\frac{\mathrm{d}\psi_{L_{\mathrm{d}s}}(u)}{\mathrm{d}u}\right|_{u=0} = \lambda\left[\frac{p}{\eta_1} - \frac{q}{\eta_2}\right]\mathrm{d}t \tag{C.14}$$

When we substitute this into the expectation under $\mathcal{F}_t$, we obtain:

$$\begin{aligned}\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)}dL(s) \mid \mathcal{F}_t\right] &= \int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}\mathbb{E}\left[L(\mathrm{d}s) \mid \mathcal{F}_t\right] \\ &= \lambda\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)\int_t^T \mathrm{e}^{-\beta(T-s)}\mathrm{d}s \\ &= \frac{\lambda}{\beta}\left(\frac{p}{\eta_1} - \frac{q}{\eta_2}\right)\left(1 - e^{-\beta(T-t)}\right)\end{aligned} \tag{C.15}$$

For the expected value under the filtration $\mathcal{G}_t$, we encourage the reader to consult section 4.3 which relies heavily on theorem 2.2.1 and corollary 2.2.1, [2]. The result is:

$$\mathbb{E}\left[\int_t^T \mathrm{e}^{-\beta(T-s)}dL(s) \mid \mathcal{G}_t\right] = \frac{1}{\beta}\left(\frac{\mathbb{E}\left[L_{T_\Upsilon} - L_t \mid \mathcal{G}_t\right]}{T_\Upsilon - t}\right)\left(1 - e^{-\beta(T-t)}\right) \tag{C.16}$$

hence, in the end, we obtain the following expression for the information premium for some knowledge about the spikes in the model:

$$I_{\mathcal{G}}^{\mathbb{P}}\left(t, T; T_\Upsilon\right) = \frac{1}{\beta}\left(\frac{\mathbb{E}\left[L_{T_\Upsilon} - L_t \mid \mathcal{G}_t\right]}{T_\Upsilon - t} + \lambda\left[\frac{p}{\eta_1} - \frac{q}{\eta_2}\right]\right)\left(1 - e^{-\beta(T-t)}\right) \tag{C.17}$$

D Additional Figures for Statistical Analysis on the Short Period

D.1 The Simple Model in eq. (E.3.1)

The parameters are:

	2.5 %	β	97.5 %
β	0.0064	0.0080	0.0097
α_1	22.75	24.37	25.98
α_2	25.84	27.45	29.05
α_3	24.88	26.49	28.09
α_4	24.38	25.99	27.60
α_5	22.15	23.76	25.37
α_6	15.64	17.25	18.86
α_7	10.48	12.09	13.70
σ^2	65.97	71.71	78.24

Table D.1 – Parameter estimates for eq. (E.3.1).

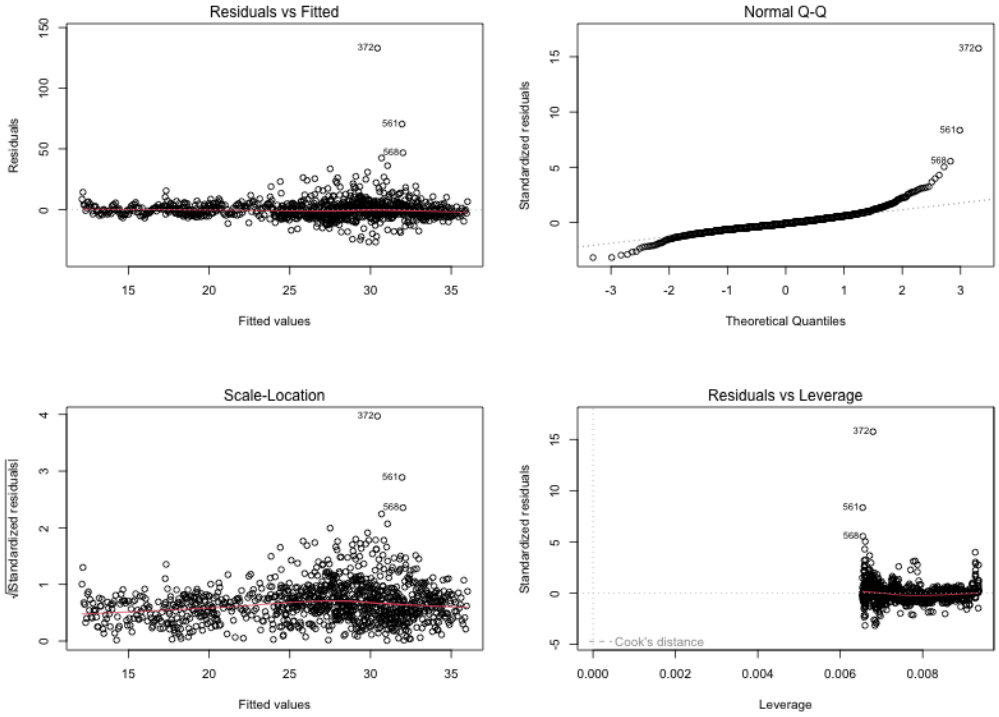


Figure D.1 – The full residual diag. plots.

D.2 The Simple Model in eq. (E.3.2)

	2.5 %	β	97.5 %
β	0.000310	0.000367	0.000424
α_1	3.062	3.118	3.173
α_2	3.141	3.196	3.251
α_3	3.129	3.184	3.239
α_4	3.103	3.158	3.213
α_5	3.050	3.105	3.160
α_6	2.790	2.845	2.900
α_7	2.511	2.566	2.621
σ^2	0.0771	0.0838	0.0914

Table D.2 – Parameter estimates for eq. (E.3.2).

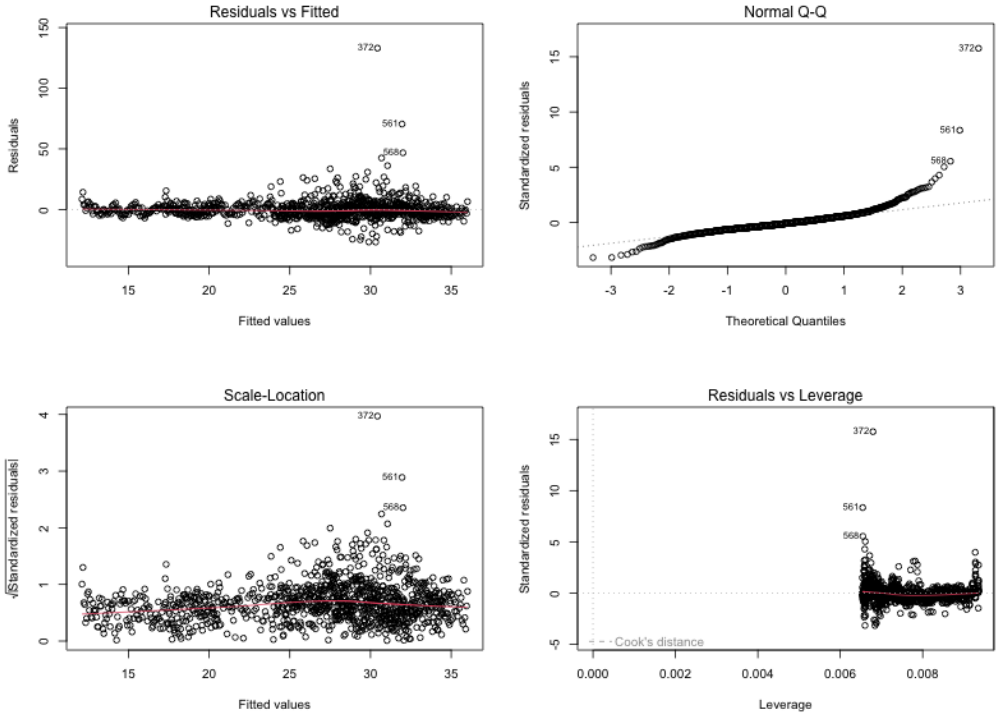


Figure D.2 – The full residual diagnostics for the model in eq. (E.3.2).

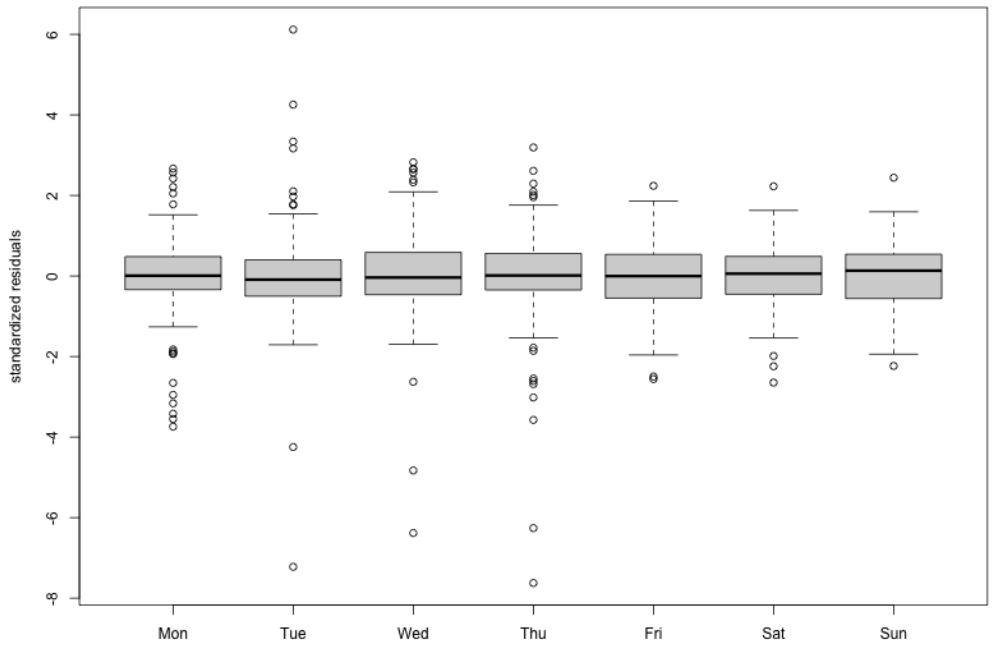


Figure D.3 – Residuals for each weekday to see spikes for the model in eq. (E.3.2).

D.3 Inclusion of Different Trends for each Weekday

To describe more of the systematic variance, we see in fig. D.2, we try to expand to an intercept for each weekday i.e. if we use the **R** syntax also used in [23], we fit the model:

$$\log \texttt{price} \sim \texttt{t} * \texttt{weekday}. \tag{D.1}$$

	2.5 %	β	97.5 %
t	0.000219	0.000369	0.000519
factor(weekday)Mon	3.039474	3.132077	3.224679
factor(weekday)Tue	3.130139	3.221663	3.313188
factor(weekday)Wed	3.108388	3.200340	3.292292
factor(weekday)Thu	3.080030	3.172112	3.264194
factor(weekday)Fri	3.012007	3.104219	3.196431
factor(weekday)Sat	2.740157	2.832500	2.924842
factor(weekday)Sun	2.415867	2.508339	2.600812
t:factor(weekday)Mon	-0.000242	-0.000029	0.000183
t:factor(weekday)Sat	-0.000191	0.000021	0.000234
t:factor(weekday)Sun	-0.000106	0.000107	0.000319
t:factor(weekday)Thu	-0.000241	-0.000028	0.000184
t:factor(weekday)Tue	-0.000262	-0.000051	0.000161
t:factor(weekday)Wed	-0.000244	-0.000032	0.000181
σ^2	0.0773	0.0840	0.0917

Table D.3 – Parameter estimates for eq. (D.1).

As one can see in fig. D.4, there is no major improvements in the diagnostics, and likewise when we do a partial likelihood ratio test as defined in definition, [23], we obtain the following table that suggests that we cannot reject the null-hypothesis that there is no difference between the models:

f	Sum of Squares	Test statistic, F
p-value		
6	0.245	0.486
0.819		

Table D.4 – Partial Likelihood Ratio Test.

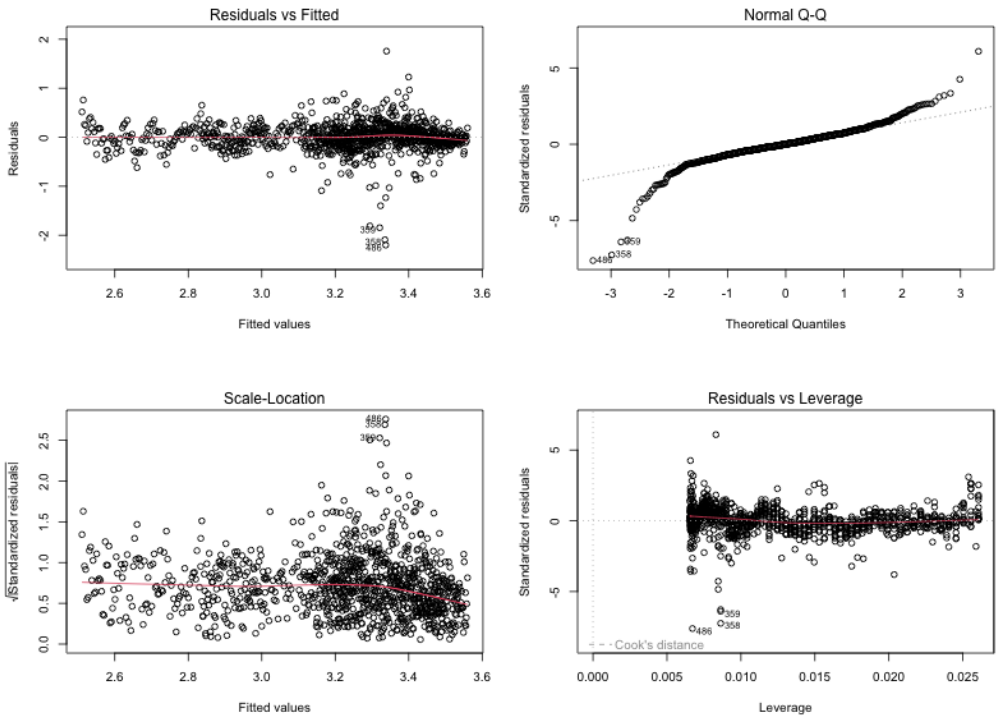


Figure D.4 – The full residual diagnostics plots with trend for each weekday

D.4 Inclusion of Different Variance for each Weekday

Here we will fit the model introduced in eq. (D.2) which is also seen below:

$$\log\left(Y_{t_i,j}\right) = \beta t_i + \alpha_1 \mathbb{1}_{\{mon\}} + \alpha_2 \mathbb{1}_{\{tues\}} + \cdots + \alpha_7 \mathbb{1}_{\{sun\}} + \epsilon_{t_i,j}. \tag{D.2}$$

As stipulated before, we will have a different variance parameter for each weekday, however, the exact implementation has some technicalities described in e.g. section 5.2 [25] under `varIdent`. We have a shared σ and then we introduced the following variance parameters, δ.

$$\epsilon_{\cdot,j} \sim \mathcal{N}\left(0, \sigma^2\delta_j^2\right), \qquad j \in \{mon, tue, \ldots, sun\} \tag{D.3}$$

In this setup, we have 8 variance parameters to represent 7 variances, and therefore to ensure identifiability, we need to impose a restriction the restriction that $\delta_{weekday_1}) = 1$ and then the remaining δ_j's would be the ratio between the standard deviation of $\delta_{weekday_1}$) and the standard deviation of weekday j. When we do that, we obtain the following parameters and variance parameters found using *restricted maximum likelihood*, REML.

	2.5 %	β	97.5 %
β	0.000327	0.000369	0.000412
α_1	3.104	3.145	3.185
α_2	3.144	3.184	3.223
α_3	3.148	3.189	3.231
α_4	3.160	3.202	3.244
α_5	3.066	3.109	3.152
α_6	2.809	2.849	2.889
α_7	2.523	2.565	2.607
σ^2		0.03987	
δ_{tue}		1.000	
δ_{wed}		1.057	
δ_{thu}		1.080	
δ_{fri}		1.162	
δ_{sat}		1.040	
δ_{sun}		1.129	
δ_{mon}		1.023	

Table D.5 – Parameter estimates for eq. (D.2).

If we re-estimate using ML, we can again make a partial likelihood ratio test and here we find that we cannot reject the null hypothesis:

df	Test statistic, F	p-value
6	5.19	0.52

Table D.6 – Partial Likelihood Ratio Test.

D.5 Inclusion of AR(1)-correlation structure for each Weekday

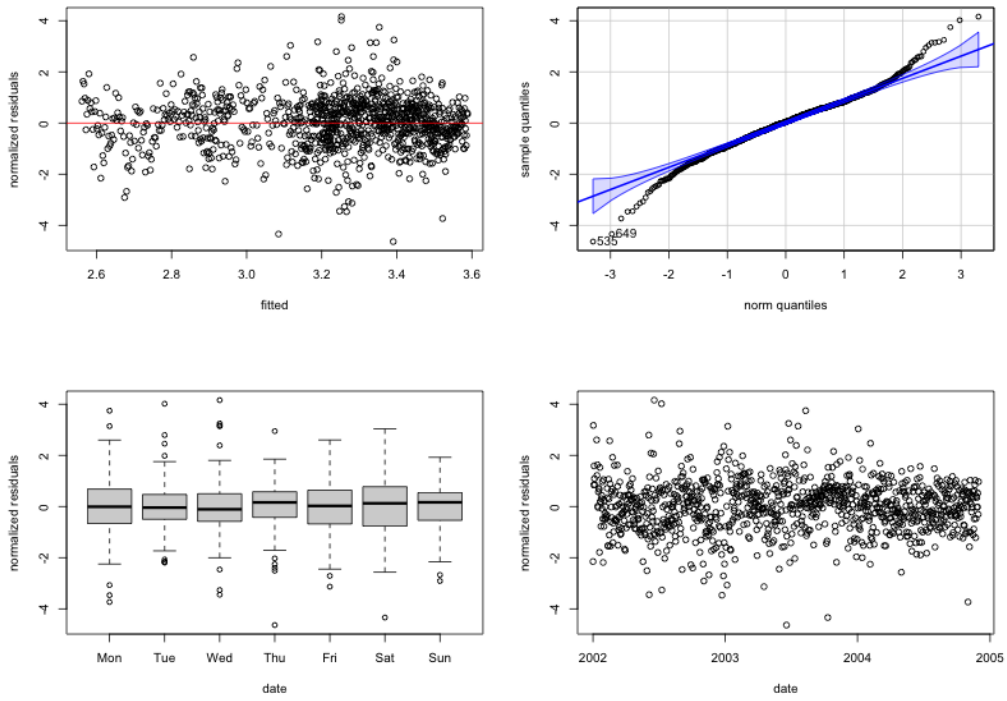


Figure D.5 – More diagnostic plots for the model with AR(1) variance structure.

df	Test statistic, F	p-value
1	606.40	<0.001

Table D.7 – Partial Likelihood Ratio Test.

E Analysis for the Entire Period

In the following, we will briefly show the analysis we did on the entire dataset and the problem that occurred.

E.1 The Seasons in Our Application

Multiple Seasonal-Trend decomposition using Loes

We will instead use MSTL [28], *Multiple Seasonal-Trend decomposition using Loes*, which is an additive decomposition method for multiple seasons. It is built on the framework of STL [29], *seasonal-trend decomposition*. Consider a series of observations $\{Z_t\}$ for $t \in \{t_0, t_1, \ldots, t_N\}$, then we can define the decomposition as:

$$Z_t = S_t^1 + S_t^2 + \cdots + S_t^{n_s} + T_t + R_t \tag{E.1}$$

where S_t^1 is the first seasonal component, S_t^2 is the second seasonal, S_t^2 is the n_s'th seasonal, T_t is the trend component, and R_t is the remainder. In case of $n_s = 1$, MSTL simplifies

and is STL. However, in our case, this decomposition is useful as we have at least a yearly and weekly seasonal component.
Each seasonal component and trend component is estimated iteratively and smoothed with loess. We will not cover the full process in detail but cover the essentials step-by-step.
In the following, we will consider as in our case a series of daily data and we want to find a weekly and yearly seasonal component. First, start by estimating the trend function directly using simple least squares and denote it $\bar{T}_t$. Then follow the steps:

1. *detrending* Calculate the detrended series by calculating $Y_t - \bar{T}_t$
2. *calculate the first seasonal signal* consider first the weekly seasonality. Let $n_p = n_7$ denote the number of data points between each period and now construct 7 *cycle subseries* of detrended data e.g. in our case a subseries of all Mondays. For each cycle subseries in the data, find the smoothed estimate with loess and $q = n_s$ and denote this C_t^1. To prevent low-frequency power from entering the seasonal component, low-pass filtering of C_t^1 is performed to obtain S_t^1, see [29] for further details.
3. *calculate the remaining seasonal signal* take the series $Y_t - S_t^i$ and repeat from step 1 to obtain season S_t^{i+1}. Do this for $i = 1, 2, \ldots, n_{s-1}$ such that we have $S_t^1, S_t^2, \ldots, S_t^s$.
4. *deseasonalize* Calculate the deseasonalized series as $Y_t - \sum_{i=1}^{n_s} S_t^i$.
5. *trend smoothing* The deseasonalized series is now smoothed by loess with $q = n_{trend}$ to obtain T_t.

In [29] [28] there are more rigors introductions and additional details on how these steps can be repeated to obtain finer estimates. The main idea is to set $\bar{T}_t = T_t$ and go through each step again to get a more refined estimate.
These steps should be integrated with the steps to find the spikes Y_t and where we rely on [2].
For our application, we choose 3 seasons, *weekly, semi-annual* and *annual*, and include also a trend for the entire dataset.

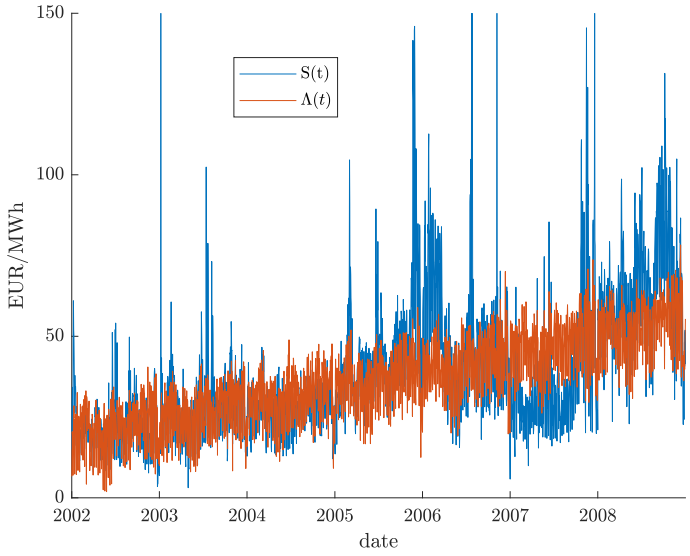
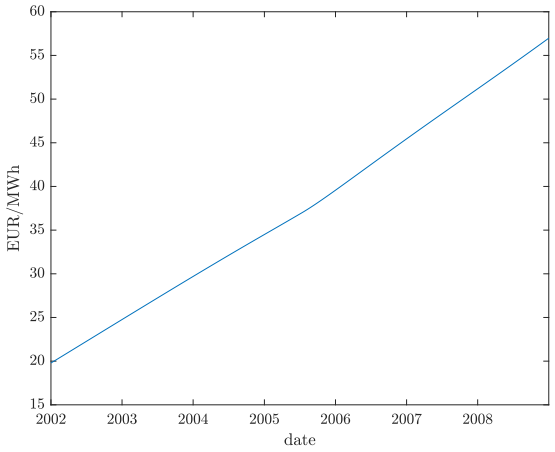
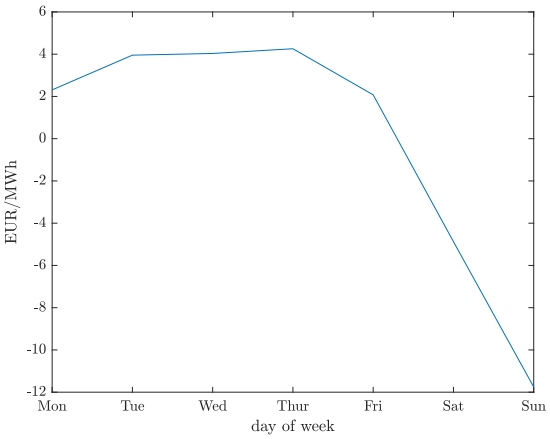


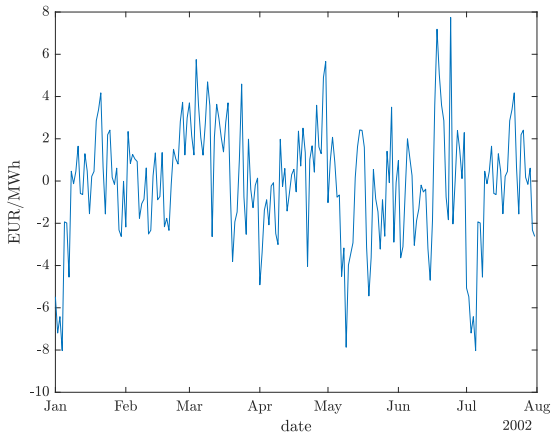
Figure E.1 – The deterministic function with the spot prices in the long period.



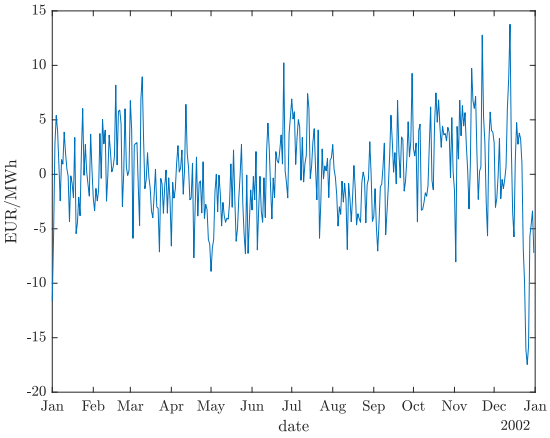
(a) Trend Component of the Deterministic Model.



(b) Weekly Component of the Deterministic Model.



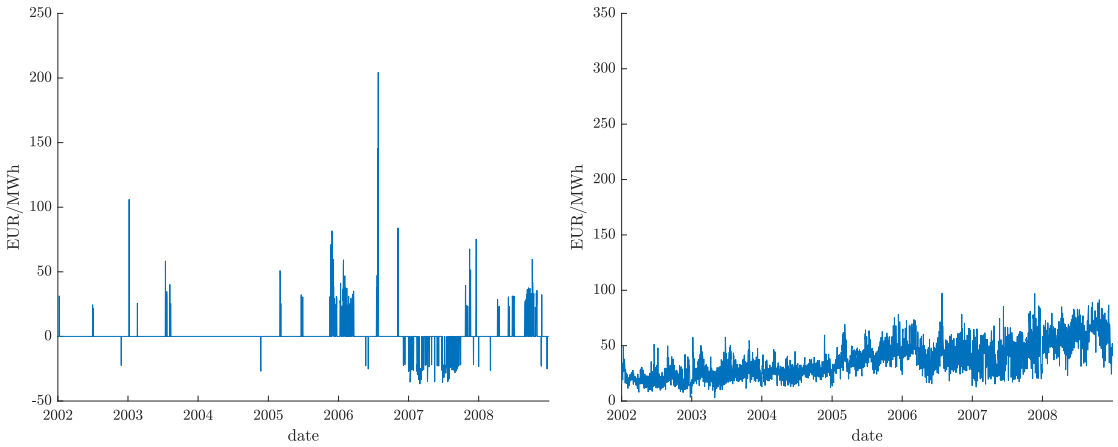
(c) Semi-annual Component of the Deterministic Model.



(d) Annual Component of the Deterministic Model.

As we can see in the figures above, the trend and the weekly component seem to make sense but the semi-annual and annual components do not seem to be as deterministic and it is hard to see and clear decomposition. We can then use the same approach as in [2] to filter the spikes. When we do that, we obtain:

E.2 The Spike Function Y_t



(a) The Pure Spike Signal Y_t

(b) Spikes Removed from Spot Price, $S_t - Y_t$.

We see that the spikes seem to be filtered when we now see the spot price without the spikes in fig. E.3b. However, we see that the spikes are so clustered fig. E.3a and they seem to cluster just around the periods where the general dynamics in the spot are shifting due to the macro economical factors explained in section E.1.1. This is already a sign that we have a poor model. From here, we could proceed and try to estimate the OU parameters as we will do in the following.

E.3 The Ornstein Uhlenbeck Process in the Signal

The rest of the signal we now assume will be determined by an Ornstein Uhlenbeck Process.

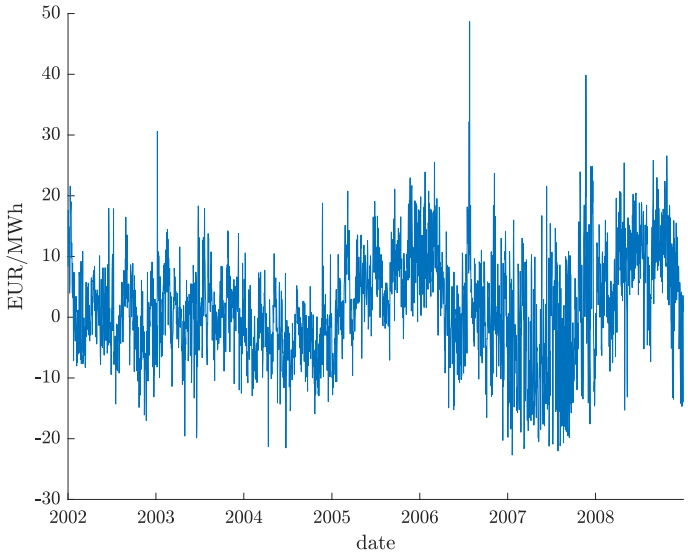


Figure E.4 – The signal to be estimated by a Brownian $S_t - (\Lambda(t) + Y_t)$

In fig. E.4, we see that this is far from a OU-process. There is lots of systematic variances from 2004 and upwards. We can compare this directly to the spot price and see that this systematic variance is seen exactly where we have the described macro-economic events and factors, revisit section E.1.1. We, therefore, chose to end the analysis here and consider

only the period from 2002 to 2004 as in the main report. To describe this period one would need a more elaborate model that included the macro economical information.