A multifactor approach to modelling the impact of wind energy on electricity spot prices

Paulina A. Rowińska[a], Almut E.D. Veraart[b], and Pierre Gruet[c]

[a] *p.rowinska15@imperial.ac.uk*

[b] *a.veraart@imperial.ac.uk*

[c] *pierre.gruet@edf.fr*

[a,b] *Department of Mathematics, Imperial College London, 180 Queen's Gate, SW7 2AZ London, United Kingdom*

[c] *Département Osiris, EDF Research and Development, 7 Boulevard Gaspard Monge, 91120 Palaiseau, France*

Abstract

We introduce a three-factor model of electricity spot prices, consisting of a deterministic seasonality and trend function as well as short- and long-term stochastic components, and derive a formula for futures prices. The long-term component is modelled as a Lévy process with increments belonging to the class of generalised hyperbolic distributions. We describe the short-term factor by Lévy semistationary processes: we start from a CARMA(2,1), i.e. a continous-time ARMA model, and generalise it by adding a short-memory stochastic volatility. We further modify the model by including the information about the wind energy production as an exogenous variable. We fit our models to German and Austrian data including spot and futures prices as well as the wind energy production and total load data. Empirical studies reveal that taking into account the impact of the wind energy generation on the prices improves the goodness of fit.

Keywords: CARMA model, Electricity spot prices, Electricity futures prices, Lévy process, Lévy semistationary process, Wind energy

JEL Classification Codes: C0, C1, C3, C5, Q4

1 Introduction

One of the main challenges of the 21^{st} century is reinforcing sustainable economic growth in order to tackle climate change. An important part of this task is a more effective use of renewable energy

sources, such as the wind power. From the economical point of view, these sources are notorious for being risky to invest in because of their unpredictable influence on the electricity prices. This is due to their high dependence on the weather – and weather forecasts still do not reach the desirable level of accuracy.

As energy markets were liberalised only in last decades, modelling electricity prices is a relatively new topic in mathematics and economics. Electricity is considered a commodity with unique futures ([Haar, 2010]) that make the use of standard tools of financial mathematics difficult or even impossible. First, the storage of large volumes of energy to date is either impossible or very expensive, so supply and demand must match at all times. Over time the prices tend to a long-term average determined by this balance, so they exhibit a mean reversion. Any disturbances of this equilibrium can result in significant spikes in the electricity spot (day-ahead) market, which leads to a strong and heteroscedastic (time-varying) volatility of electricity (especially spot) prices. Furthermore, electricity prices are seasonal. The demand is much higher in winter months (due to the need of heating and longer use of lights) as well as during hot summer months (due to the use of air conditioning). Because prices are inelastic of demand, very cold or very warm weather usually results in more expensive electricity. The periodic behaviour can also be observed at a smaller, weekly scale, namely the demand is higher in the peak time, i.e. Monday to Friday between 8 am and 8 pm, when people need electricity for their activities at work and home.

In the literature one can find a variety of electricity prices models: for spot and futures via a spot model (e.g. [Carmona et al., 2013], [Cartea et al., 2009], [Benth et al., 2014]) and for futures directly (e.g. [Benth and Paraschiv, 2016], [Barndorff-Nielsen et al., 2011], [Borovkova and Geman, 2006]). However, most of them do not take into account the increasing role of renewables, in particular wind power, in energy markets. Few exeptions include [Elberg and Hagspiel, 2015] (copula model for the spatial dependence structure of wind power in Germany), [Veraart, 2016] (impact of wind power generation on German spot prices modelled by regime-switching Lévy semistationary processes), [Ketterer, 2014] (GARCH model of wind power's impact on the electricity price level and volatility in Germany, taking into account changes in market regulations) or [Deschatre and Veraart, 2017] (the impact of wind energy production on the spikes in the spot prices). In this paper we attempt to fill this research gap by introducing a model for both spot and futures prices with wind energy production as an exogenous variable. Our work builds upon the paper by [Benth et al., 2014], who proposed an arithmetic model for spot prices with three factors: a deterministic seasonality and trend function as well as short- and long-term stochastic parts (details in Section 2). We generalised this model by: (a) considering a short-term process more general than CARMA(2,1) proposed by [Benth et al., 2014], i.e. a Lévy semistationary (LSS) process ([Barndorff-Nielsen et al., 2013]) with a stochastic volatility; (b) modelling the dependence of the short-term process on the wind energy production. We fitted the model to spot and futures data from the European Energy Exchange

(EEX) as well as the wind production data from Germany and Austria. Finally, we defined a moment-based metric to compare different models and found that the inclusion of wind energy production data improved the model fit.

The paper is structured as follows. In Section 2 we introduce a three-factor arithmetic model for spot prices and derive a formula for futures prices. In Section 3 we describe how we fitted the model to empirical data and study all model terms in detail. We also present numerical results. In the last subsection of Section 3 we discuss the necessity of adding stochastic volatility to the model. Section 4 includes the description of wind production data, possible modifications of the basic model, numerical results and comparison of different models. Finally in Section 5 we present the conclusions of our research.

2 The arithmetic model

2.1 Assumptions

We work on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{R}}, \mathbb{P})$ satisfying the usual conditions, i.e. the right-continuity of the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}}$ and completeness. Let $S(t)$ be the spot price. Following [Benth et al., 2014], we propose an arithmetic model $S(t) = \Lambda(t) + Z(t) + Y(t)$, where $Z(t)$ is the long-term factor, while $Y(t)$ describes the short-term behaviour. Precisely, $\Lambda(t)$ denotes a deterministic seasonality and trend function, $Z(t)$ is a Lévy process with zero mean (under the physical measure) and $Y(t) = \int_{-\infty}^{t} g(t-s)\sigma_{s-}dL_s$ with a deterministic kernel $g(t-s)$ such that $\lim_{t\to\infty} g(t-s) = 0$. For the integrability conditions we refer the reader to Appendix A. The short-term process belongs to a class of Lévy semistationary (LSS) processes ([Barndorff-Nielsen et al., 2013]), so $Y(t)$ is stationary if and only if σ_t and the increments of $L(t)$ are jointly stationary. Note that integrating from $-\infty$ does not correspond to the real world, where $t \geq 0$. However, we need such an integral to obtain a stationary model. Therefore, similarily to [Barndorff-Nielsen et al., 2013], we assume that $S(0)$ is a realisation of a random variable $\Lambda(0)+Z(0)+Y(0)$. Here σ_t is a càdlàg stochastic process describing the volatility of $Y(t)$. The empirical studies (Section 3) indicate that σ_t has a rather short memory. Therefore we can restrict ourselves to a (stationary) Ornstein-Uhlenbeck process, i.e. $\sigma_t = \int_{-\infty}^{t} e^{-\delta(t-x)}dV_x$ with a constant $\delta > 0$ and a Lévy subordinator $V(t)$, independent from the driving Lévy process $L(t)$.

2.2 Change of measure

In the traditional finance theory, one requires a probability measure $\mathbb{Q}$ equivalent to the physical measure $\mathbb{P}$ that transforms the discounted price dynamics into a (local) $\mathbb{Q}$-martingale. However, in the electricity markets not all the assets are tradeable, so $\mathbb{Q}$ can denote any probability measure

equivalent to $\mathbb{P}$. As suggested by [Barndorff-Nielsen et al., 2013], we decided to use the generalised Esscher transforms; details are described in Appendix C.

2.3 Futures contracts

If we assume no arbitrage, the price $f_t(T)$ at the time $t \geq 0$ for a futures contract with maturity $T \geq t$ can be expressed as

$$\begin{aligned} f_t(T) &= \mathbb{E}_{\mathbb{Q}}\left[S(T)|\mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}[\Lambda(T) + Z(T) + Y(T)|\mathcal{F}_t] = \Lambda(T) + Z(t) + (T-t)\mathbb{E}_{\mathbb{Q}}\left[Z(1)\right] \\ &+ \int_{-\infty}^{t} g(T-s)\sigma_{s-}dL_s + \mathbb{E}_{\mathbb{Q}}\left[L_1\right]\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds. \end{aligned} \tag{2.1}$$

Due to its nonstorability electricity is delivered over a time period rather than at one specific moment. Thus we define the price of a futures contract with a delivery period $[T_1, T_2]$ as

$$F_t(T_1, T_2) := \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{T_2 - T_1}\int_{T_1}^{T_2} S(T)dT\middle|\mathcal{F}_t\right], \tag{2.2}$$

for all $0 \leq t \leq T_1 < T_2$. If following [Benth et al., 2014] we define time to maturity as $u := \frac{1}{2}(T_1 + T_2) - t$, then Equation 2.2 becomes

$$\begin{aligned} F_t(T_1, T_2) &= \frac{1}{T_2 - T_1}\int_{T_1}^{T_2}\Lambda(T)dT + Z(t) + u\mathbb{E}_{\mathbb{Q}}\left[Z(1)\right] \\ &+ \frac{1}{T_2 - T_1}\left(\int_{T_1}^{T_2}\int_{-\infty}^{t} g(T-s)\sigma_{s-}\,dL_s\,dT + \mathbb{E}_{\mathbb{Q}}\left[L_1\right]\int_{T_1}^{T_2}\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]\,ds\,dT\right). \end{aligned} \tag{2.3}$$

Proposition 2.1. *Assume that* $\delta > 0$, $\lim_{x\to\infty}\int_0^x g(y)e^{-\frac{\delta}{2}(x-y)}dy = 0$ *and* $\sigma_t^2 = \int_{-\infty}^t e^{-\delta(t-x)}dV_x$. *Then for* $\tau > 0$ *and fixed* $t > 0$,

$$\begin{aligned} &\lim_{T_1\to\infty}\frac{1}{\tau}\left(\int_{T_1}^{T_1+\tau}\int_{-\infty}^{t} g(T-s)\sigma_{s-}\,dL_s\,dT + \mathbb{E}_{\mathbb{Q}}\left[L_1\right]\int_{T_1}^{T_1+\tau}\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]\,ds\,dT\right) \\ &= \mathbb{E}_{\mathbb{Q}}\left[L_1\right]C, \end{aligned} \tag{2.4}$$

where

$$C := \mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]\int_0^\infty g(y)dy \tag{2.5}$$

and the limit is in the L^2-*sense.*

For the proof we refer the reader to Appendix B. Proposition 2.1 allows us to conclude that in the long end, i.e. for $t \ll T_1$, the deseasonalised futures price can be approximated by

$$\tilde{F}_t(T_1, T_2) := F_t(T_1, T_2) - \frac{1}{T_2 - T_1}\int_{T_1}^{T_2}\Lambda(T)dT \approx Z(t) + u\mathbb{E}_{\mathbb{Q}}\left[Z(1)\right] + \mathbb{E}_{\mathbb{Q}}\left[L_1\right]C. \tag{2.6}$$

3 Empirical studies

We fitted the proposed model to two sets of German data: daily averages of hourly spot prices (available on the EPEX SPOT website [1]) and monthly (one month ahead, 1MAH, up to six months ahead, 6MAH) base load futures prices, (available on the EEX website [2]), traded between 1 January 2013 and 3 August 2015. In order to split spot prices into $S(t) = \Lambda(t) + Z(t) + Y(t)$, we modified the algorithm proposed by [Benth et al., 2014, pp. 398-9]. It requires a choice of threshold $u^\star$ such that deseasonalised prices of contracts with times to maturity $u \geq u^\star$ are described approximately by $Z(t)$ alone; recall that we defined 'time to maturity' as $u = \frac{1}{2}(T_1 + T_2) - t$. We decided to fix $u^\star$ equal to 16 days, because the results do not seem to be sensitive to the choice of a particular threshold. The algorithm can be summarised as follows.

1. Estimate $\Lambda(\cdot)$ from spot prices and subtract from $S(\cdot)$ (Subsection 3.1).
2. Filter out a realisation of $Z(t)$ (Subsection 3.2).
3. Model $Y(t) = S(t) - \Lambda(t) - Z(t)$ (Subsection 3.3) as a CARMA(2,1) process.
4. Add stochastic volatility to the model of $Y(t)$ (Subsection 3.4).

3.1 Deterministic seasonality and trend function $\Lambda(t)$

The exact form of the seasonality function is not obvious and in the literature it appears in different forms, usually as a combination of trigonometric functions (see e.g. [Benth et al., 2014]) or polynomials. We decided to use the following function, which works well with our data set:

$$\Lambda(t) = c_1 + c_2 t + c_3 h(t) + \sum_{i=4}^{9} c_i d(t) + \sum_{i=10}^{13} c_i m(t). \tag{3.1}$$

It consists of a linear trend as well as dummy variables for days of the week $d(t)$ and statistically significant months $m(t)$: April, May, June and August. We also took into account rapid downward movements of prices during the main holidays, Christmas (24, 25, 26 December) and New Year (1 January), by introducing a dummy variable

$$h(t) = \begin{cases} 0 & \text{for holidays,} \\ 1 & \text{otherwise.} \end{cases} \tag{3.2}$$

We fitted $\Lambda(t)$ using linear regression. The estimated coefficients are presented in Table 1. Figure 1 shows the autocorrelation plot of deseasonalised spot prices. It seems that we managed to remove most of the periodic behaviour, although some seasonal effects might be still present.

[1] `https://www.epexspot.com/en/market-data/dayaheadauction` (accessed 10.11.2017)
[2] `https://www.eex.com/en/market-data/power/futures` (accessed 10.11.2017)

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	5.3119	2.6300	2.02	0.0437
t	-0.0095	0.0009	-10.59	0.0000
Holiday	22.8892	2.5195	9.08	0.0000
Apr	-3.4478	0.8444	-4.08	0.0000
May	-6.5892	0.8332	-7.91	0.0000
Jun	-5.6228	0.8465	-6.64	0.0000
Aug	-3.5180	0.9759	-3.60	0.0003
Mon	14.1859	0.9089	15.61	0.0000
Tue	17.0706	0.9098	18.76	0.0000
Wed	17.2351	0.9108	18.92	0.0000
Thu	16.7177	0.9107	18.36	0.0000
Fri	15.0050	0.9092	16.50	0.0000
Sat	7.3066	0.9089	8.04	0.0000

Table 1: Estimated coefficients of the seasonality function $\Lambda(\cdot)$.

3.2 Non-stationary long-term factor $Z(t)$

Let us denote the empirical mean (i.e. the averaged data) by $\widehat{\mathbb{E}}[\cdot]$. For $u \geq u^\star$, using Proposition 2.1 and $\mathbb{E}\left[Z(t)\right] = 0$, we can approximate

$$\mu_{\tilde{F}}(u) := \widehat{\mathbb{E}}\left[\tilde{F}_t(T_1, T_2)\right] = u\widehat{\mathbb{E}}_{\mathbb{Q}}\left[Z(1)\right] + \widehat{\mathbb{E}}_{\mathbb{Q}}\left[L_1\right] C \tag{3.3}$$

and by linear regression estimate $\widehat{\mathbb{E}}_{\mathbb{Q}}\left[L_1\right] C$ and $\widehat{\mathbb{E}}_{\mathbb{Q}}\left[Z(1)\right]$, as presented in Figure 2. These parameters together allow us to recover a realisation of $Z(t)$:

$$\widehat{Z}(t) = \widehat{Z}\left(\frac{1}{2}(T_1 + T_2) - u\right) = \frac{1}{\mathrm{card}U(t, u^\star)} \sum_{(u, T_1, T_2) \in U(t, u^\star)} \left[\tilde{F}_t(T_1, T_2) - \widehat{\mathbb{E}}_{\mathbb{Q}}\left[L_1\right] \widehat{C} - u\widehat{\mathbb{E}}_{\mathbb{Q}}\left[Z(1)\right]\right], \tag{3.4}$$

where

$$U(t, u^\star) := \{(u, T_1, T_2) \in \mathbb{R}^3 : u \geq u^\star \text{and} \quad \exists F_t(T_1, T_2) : \frac{1}{2}(T_1 + T_2) - t = u\}. \tag{3.5}$$

Estimated $Z(t)$ is plotted in Figure 3. It might seem to exhibit annual seasonality, but we fitted an ETS (Error-Trend-Seasonality) model using the R package **forecast** and the best fit does not have a seasonal component.

Because futures contracts are traded only from Monday to Friday, $Z(t)$ does not include the weekend data. Following [Benth et al., 2014], on the weekends we set this process to be constant and equal to the Friday price. The process $Z(t)$ is clearly non-stationary, therefore we focused our

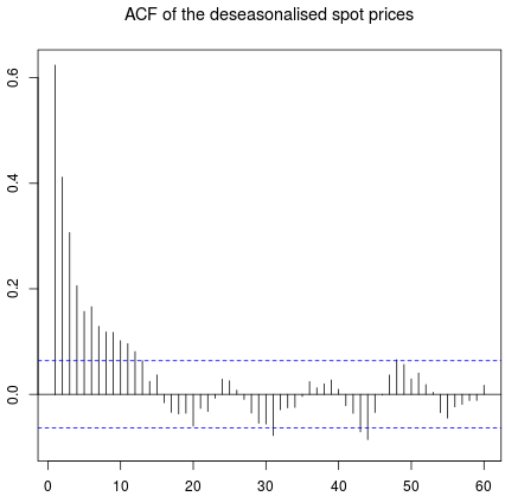

Figure 1: ACF of deseasonalised spot prices.

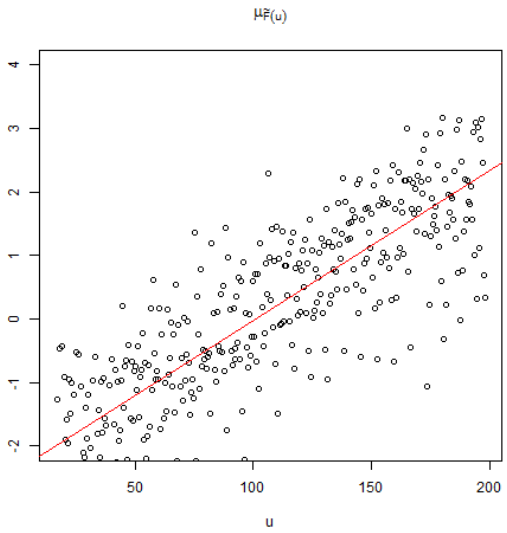

Figure 2: Plot of $\mu_{\tilde{F}}(u)$.

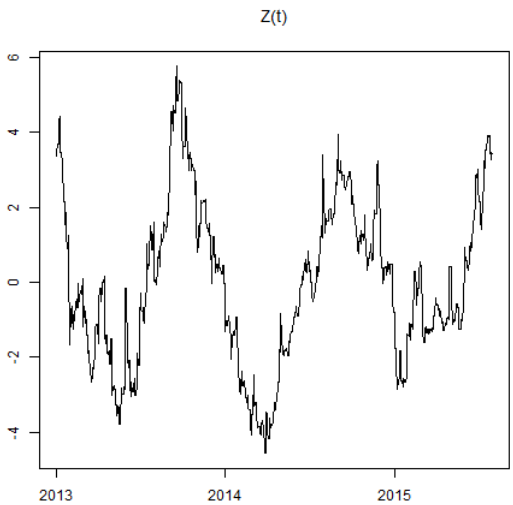

Figure 3: Estimated $Z(t)$.

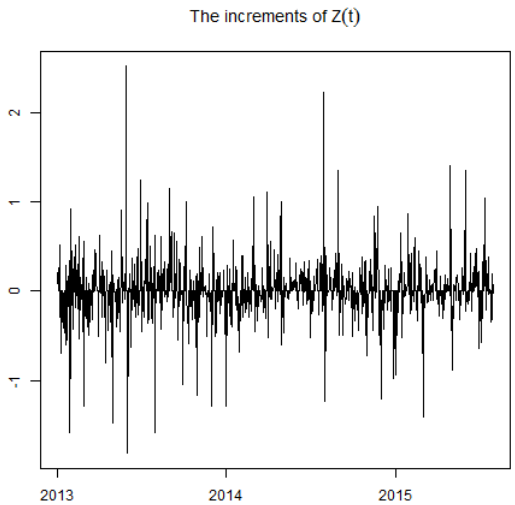

Figure 4: The increments of $Z(t)$.

attention on its increments, plotted in Figure 4. At first glance $Z(t)$ has stationary increments, as confirmed at 0.01 significance level by the augmented Dickey-Fuller (ADF) test. It is important to note that the estimated $Z(t)$ has uncorrelated increments and squared increments, as shown in Figure 5 and Figure 6. This motivated us to model $Z(t)$ by a Lévy process with increments described by a suitable infinitely divisible distribution. Inspired by [Barndorff-Nielsen et al., 2013],

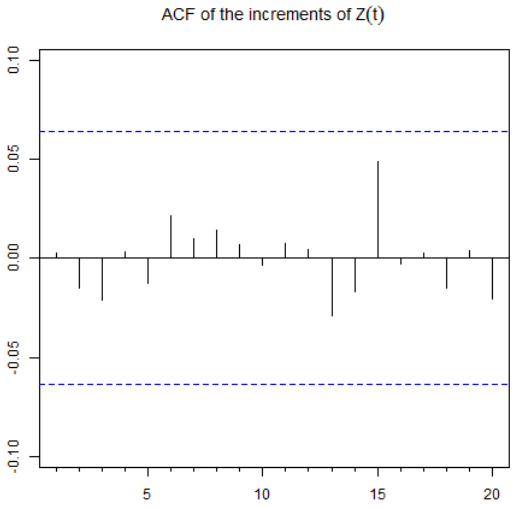

Figure 5: ACF of the increments of $Z(t)$.

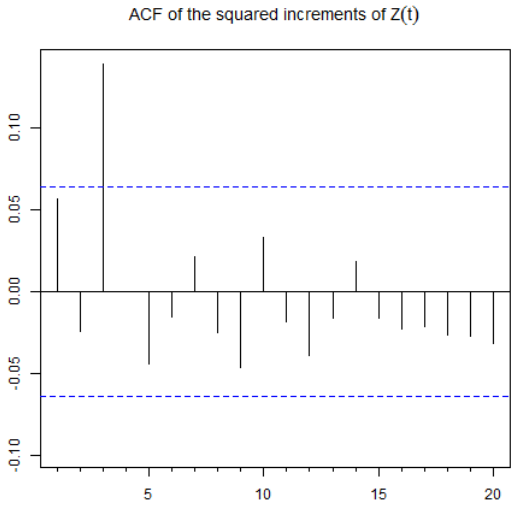

Figure 6: ACF of the squared increments of $Z(t)$.

we fitted 11 distributions from a class of generalised hyperbolic distributions, defined e.g. by [Breymann, 2011]. This class of distributions has some desirable properties, such as fat tails and skewness.

Definition 3.1. A random vector X is said to have a multivariate generalised hyperbolic (GH) distribution if

$$\mathbf{X} \overset{law}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}\mathbf{A}\mathbf{Z}, \tag{3.6}$$

where $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_k)$, $\mathbf{A} \in \mathbb{R}^{d\times k}$, $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^d$ and W is a scalar-valued random variable, independent of $\mathbf{Z}$, whose distribution is Generalised Inverse Gaussian $GIG(\lambda, \chi, \psi)$ with parameters satisfying one of the following: $\chi > 0, \psi \geq 0, \lambda < 0$ or $\chi > 0, \psi > 0, \lambda = 0$ or $\chi \geq 0, \psi > 0, \lambda > 0$. The density of W is given by

$$f_{\text{GIG}}(x) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{x^{\lambda-1}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} \exp\left(-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right), \tag{3.7}$$

where K_λ denotes the modified Bessel function of the third kind.

Observe that there exist different ways of parametrising generalised hyperbolic distributions and in our analysis we used the $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ parametrisation. One can easily switch from

$(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ to $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$ by setting $k = \sqrt{\frac{\chi}{\psi}} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}$, $\bar{\alpha} = \sqrt{\chi\psi}$, $\Sigma = k\Sigma$ and $\gamma = k\gamma$. The reparametrisation in the opposite direction can be done by setting $\psi = \bar{\alpha}\frac{K_{\lambda+1}(\bar{\alpha})}{K_\lambda(\bar{\alpha})}$ and $\chi = \frac{\bar{\alpha}^2}{\psi}$, while the remaining parameters stay the same.

We used the R package **ghyp** provided by [Breymann, 2011]. We ranked distributions according to the Akaike information criterion (AIC) in ascending order, so the first one gives the best fit. In Table 2 we can see that the increments of $Z(t)$ are best described by a symmetric Normal-inverse Gaussian (NIG) distribution. The fit is satisfactory, as shown in Figure 7. The QQ-plot presented in Figure 8 also proves a reasonably good fit. Therefore we can model $Z(t)$ by a Lévy process $\bar{Z}(t)$ such that $\bar{Z}(1) \sim NIG(\lambda = -0.500, \bar{\alpha} = 0.405, \mu = -0.002, \Sigma = 0.396, \gamma = 0.000)$.

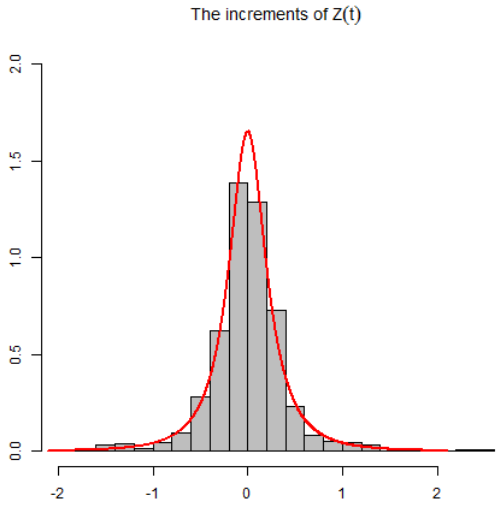

Figure 7: Histogram of the increments of $Z(t)$ with a fitted NIG distribution.

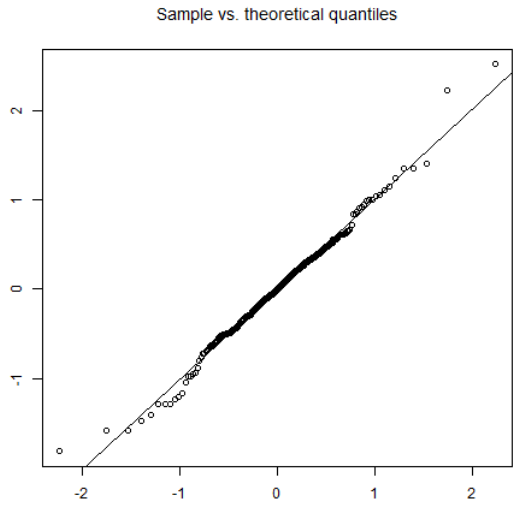

Figure 8: Symmetric generalised hyperbolic QQ-plot.

3.3 Stationary short-term factor $Y(t)$

We obtained a realisation of $Y(t)$ by subtracting $Z(t)$ from the deseasonalised spot prices. Figure 9 shows the resulting process. At first glance the plot of $Y(t)$ does not indicate if the series is stationary. However, the augmented Dickey-Fuller (ADF) test suggested that $Y(t)$ is stationary (at the significance level 0.01). Figure 10 and Figure 11 show the autocorrelation functions of $Y(t)$ and its increments, respectively.

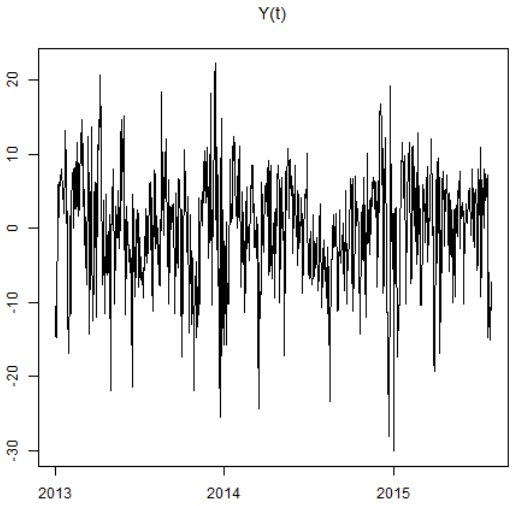

Figure 9: Estimated $Y(t)$.

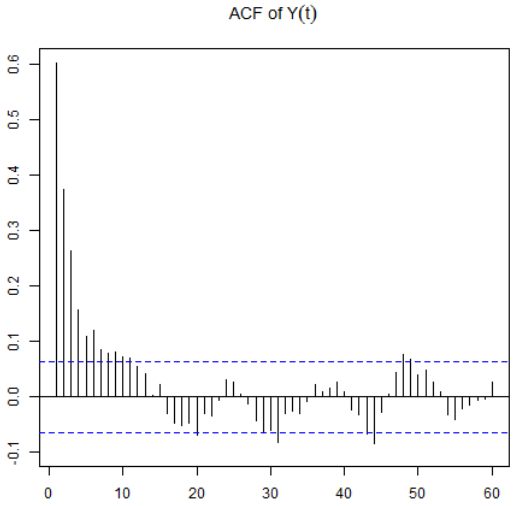

Figure 10: ACF of $Y(t)$.

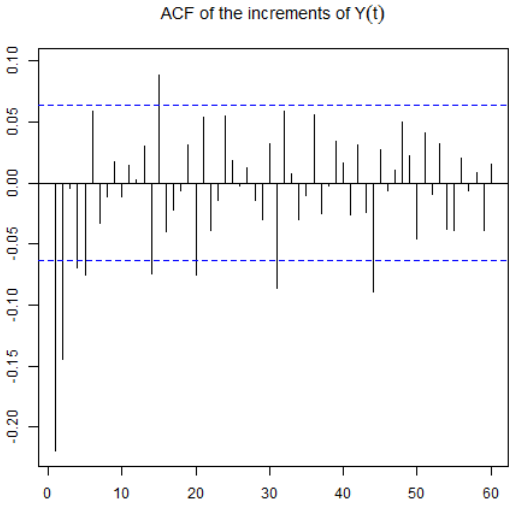

Figure 11: ACF of the increments of $Y(t)$.

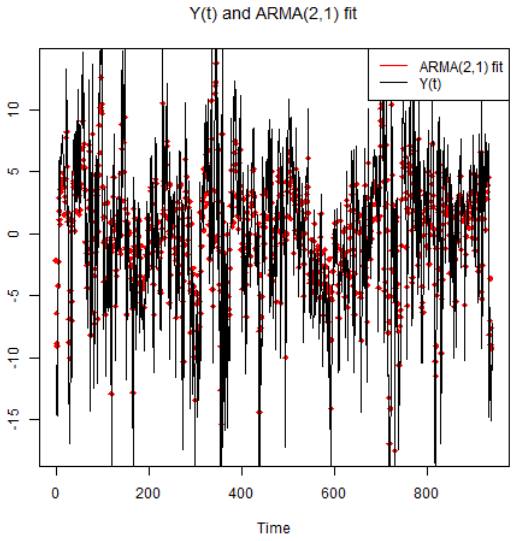

Figure 12: ARMA(2,1) fit against $Y(t)$.

	model	symmetric	λ	$\bar{\alpha}$	μ	Σ	γ	AIC	log-lik.	converged?	number of iter.
8	NIG	TRUE	-0.50	0.40	0.00	0.40	0.00	469.53	-231.76	TRUE	82.00
10	t	TRUE	-1.32	0.00	0.00	0.47	0.00	470.78	-232.39	TRUE	92.00
6	ghyp	TRUE	-0.88	0.33	0.00	0.40	0.00	470.85	-231.43	TRUE	207.00
3	NIG	FALSE	-0.50	0.41	0.00	0.40	-0.00	471.51	-231.75	TRUE	155.00
5	t	FALSE	-1.32	0.00	0.00	0.47	-0.00	472.75	-232.37	TRUE	171.00
1	ghyp	FALSE	-0.88	0.33	0.00	0.40	-0.00	472.82	-231.41	TRUE	322.00
7	hyp	TRUE	1.00	0.14	-0.00	0.37	0.00	483.31	-238.65	TRUE	98.00
9	VG	TRUE	0.95	0.00	-0.01	0.38	0.00	484.08	-239.04	TRUE	116.00
2	hyp	FALSE	1.00	0.14	-0.00	0.37	0.00	485.26	-238.63	TRUE	251.00
4	VG	FALSE	0.94	0.00	-0.01	0.38	0.01	485.76	-238.88	TRUE	185.00
11	gauss	TRUE		Inf	0.00	0.40	0.00	651.23	-323.62	TRUE	0.00

Table 2: Generalised hyperbolic distributions fitted to the increments of $Z(t)$ with parametrisation $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$.

3.3.1 CARMA processes

Inspired by [Benth et al., 2014], as a first step we decided to model $Y(t)$ by a special case of an *LSS* process, i.e. a CARMA (continuous-time ARMA) process. This class of processes was described for example by [Brockwell et al., 2011]. We focus on a particular case, i.e. the CARMA(2,1) model.

Definition 3.2. Assume that $L(t)$ is a Lévy process whose second moment is finite (i.e. $\mathbb{E}\left[L(1)^2\right] < \infty$). Furthermore, $b(z) = b_0 + z$, $a(z) = z^2 + a_1 z + a_2 = (z - \lambda_1)(z - \lambda_2)$, roots of $a(z) = 0$ are distinct and their real parts are negative as well as $a(z)$ and $b(z)$ have no common roots. The L-driven CARMA(2,1) process is defined as the strictly stationary solution to the system of stochastic differential equations

$$a(D)Y(t) = b(D)DL(t), \tag{3.8}$$

where $t \in \mathbb{R}$ and D denotes differentiation with respect to t (in a formal sense).

We can represent $Y(t)$ as

$$Y(t) = \alpha_1 \int_{-\infty}^{t} e^{\lambda_1 (t-s)} dL(s) + \alpha_2 \int_{-\infty}^{t} e^{\lambda_2 (t-s)} dL(s), \tag{3.9}$$

where $\alpha_1 = \frac{b_0 + \lambda_1}{\lambda_1 - \lambda_2}$ and $\alpha_2 = \frac{b_0 + \lambda_2}{\lambda_2 - \lambda_1}$. For $u \geq 0$ we define the kernel of $Y(t)$ as $g(u) = \alpha_1 e^{\lambda_1 u} + \alpha_2 e^{\lambda_2 u}$.

Equation 3.8 should be interpreted as its state-space representation, i.e.

$$Y(t) = \boldsymbol{b}^T \boldsymbol{X}(t), \tag{3.10}$$

$$d\boldsymbol{X}(t) = \boldsymbol{A}\boldsymbol{X}(t)dt + \boldsymbol{e}dL(t), \tag{3.11}$$

where

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -a_2 & -a_1 \end{pmatrix}, \boldsymbol{e} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_0 \\ 1 \end{pmatrix} \tag{3.12}$$

and $\boldsymbol{X}(0)$ is independent of $\{L(t), t \geq 0\}$.

Remark 3.1. Note that the CARMA(2,1) kernel satisfies the assumptions of Proposition 2.1, as long as $\lambda_i \neq -\frac{\delta}{2}$ for $i = 1, 2$. This follows from

$$\lim_{x\to\infty} \int_0^x g(y) e^{-\frac{\delta}{2}(x-y)} dy = \lim_{x\to\infty} \sum_{i=1}^{2} \alpha_i \int_0^x e^{\lambda_i y} e^{-\frac{\delta}{2}(x-y)} dy = \lim_{x\to\infty} \sum_{i=1}^{2} \frac{\alpha_i}{\lambda_i + \frac{\delta}{2}} \left(e^{\lambda_i x} - e^{-\frac{\delta}{2}x} \right) = 0, \tag{3.13}$$

as $\delta > 0$ and $\lambda_i < 0$ for $i = 1, 2$.

3.3.2 Numerical results

We implemented the algorithm from Appendix E to fit a CARMA(2,1) model to our process $Y(t)$. First we used the function *arima* from the R package **stats** to fit an ARMA(2,1) process, whose estimated parameters are shown in Table 3. In Figure 12 we can see that the estimated discrete model describes $Y(t)$ quite accurately, although it cannot capture extreme values.

	ϕ_1	ϕ_2	θ
Estimate	1.347	-0.428	-0.765
Standard error	0.196	0.127	0.188
Relative error	0.145	0.296	0.246

Table 3: Estimated ARMA(2,1) parameters.

Proceeding as described in Appendix E, we estimated CARMA(2,1) parameters presented in Table 4. The calculations of standard and relative errors as well as biases are based on the parametric bootstrapping with 1000 Monte Carlo simulations (performed using the R package **yuima**) from the estimated CARMA(2,1) process, i.e. $(D^2 + 0.847D + 0.122)Y(t) = (0.269 + D)DL(t)$. Observe that parameters were estimated with small errors, but a relatively large bias, which we further discuss in Remark 3.2.

We estimated the autoregressive roots as $\lambda_1 = -0.184$ and $\lambda_2 = -0.663$. Because $|\lambda_1| < |\lambda_2|$, we used λ_1 to recover the background driving Lévy process $L(t)$, as suggested in Appendix E. The resulting process is presented in Figure 13, while its increments in Figure 14. Figure 15 and Figure 16 indicate that the increments of $L(t)$ are not independent, so the model would probably benefit from including a stochastic volatility in the definition of $Y(t)$, at least with a short memory (see Subsection 3.4).

Similarly to the case of $Z(t)$, we fitted 11 GH distributions to the increments of $L(t)$, as presented in Table 5. The best fit was given by the asymmetric NIG distribution.

	a_1	a_2	b_0
Estimates	0.847	0.122	0.269
Standard error	0.034	0.006	0.016
Relative error	0.019	0.015	0.031
Bias	0.639	0.211	0.096

Table 4: Estimated CARMA(2,1) parameters.

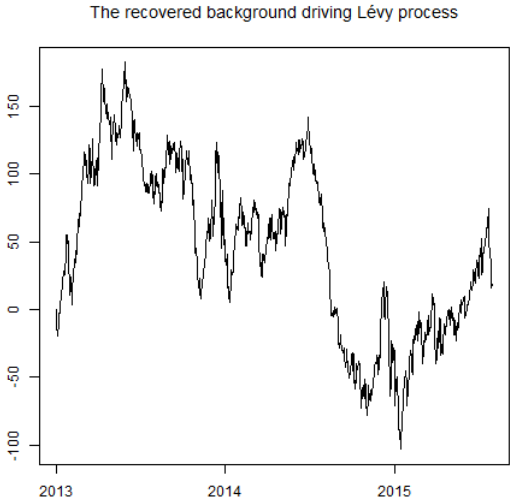

Figure 13: Background driving Lévy process $L(t)$.

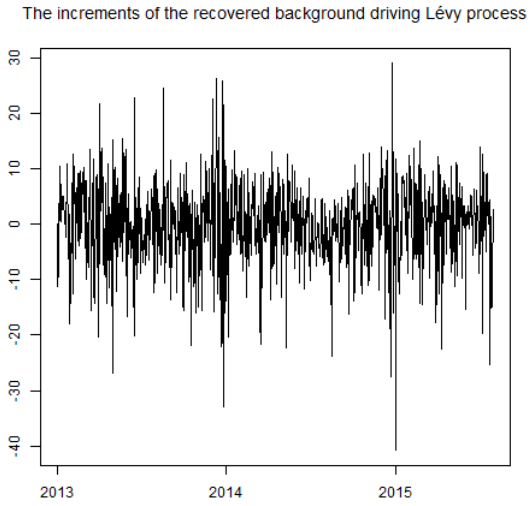

Figure 14: The increments of the background driving Lévy process $L(t)$.

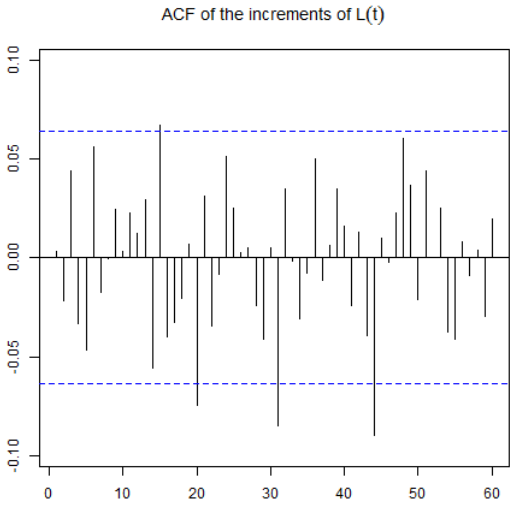

Figure 15: ACF of the increments of $L(t)$.

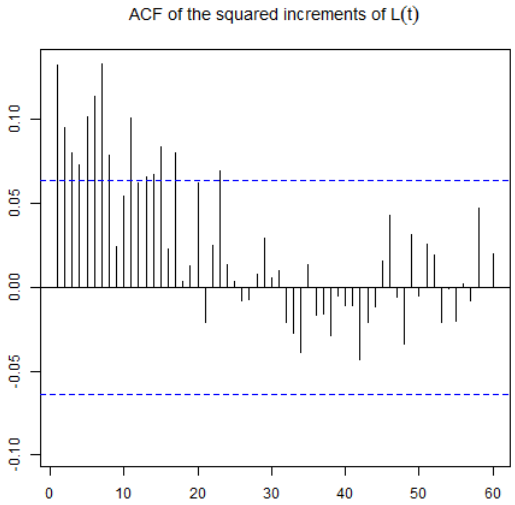

Figure 16: ACF of the squared increments of $L(t)$.

	model	symmetric	λ	$\bar{\alpha}$	μ	Σ	γ	AIC	log-lik.	converged?	number of iter.
3	NIG	FALSE	-0.50	1.96	2.81	7.26	-2.79	6413.04	-3202.52	TRUE	129.00
5	t	FALSE	-3.36	0.00	2.68	7.28	-2.67	6413.61	-3202.81	TRUE	209.00
2	hyp	FALSE	1.00	1.71	2.90	7.24	-2.88	6413.83	-3202.92	TRUE	237.00
4	VG	FALSE	2.37	0.00	2.87	7.25	-2.85	6415.10	-3203.55	TRUE	203.00
1	ghyp	FALSE	2.37	0.06	2.87	7.25	-2.85	6417.10	-3203.55	TRUE	288.00
10	t	TRUE	-3.00	0.00	0.35	7.55	0.00	6424.96	-3209.48	TRUE	82.00
8	NIG	TRUE	-0.50	1.79	0.36	7.51	0.00	6426.36	-3210.18	TRUE	84.00
6	ghyp	TRUE	-2.89	0.47	0.35	7.54	0.00	6426.96	-3209.48	TRUE	401.00
7	hyp	TRUE	1.00	1.62	0.35	7.48	0.00	6427.82	-3210.91	TRUE	104.00
9	VG	TRUE	2.37	0.00	0.36	7.48	0.00	6429.70	-3211.85	TRUE	100.00
11	gauss	TRUE		Inf	0.02	7.54	0.00	6475.14	-3235.57	TRUE	0.00

Table 5: Generalised hyperbolic distributions fitted to the increments of $L(t)$.

Remark 3.2. In our numerical studies we followed an estimation procedure of CARMA(2,1) models used in numerous studies, e.g. by [García et al., 2011] or [Brockwell et al., 2011]. However, we detected problems connected to this method which, as far as we know, have not been discussed in the literature. We decided to investigate why the bias of estimated CARMA(2,1) parameters presented in Table 4 is so large.

First, standard R functions such as *arima* from the package **stats** fit an $ARMA(2,1)$ model with a high level of variability, as shown in Figure 17 and Figure 18 (results obtained via bootstrapping). Therefore we tried to estimate the autoregressive roots λ_1 and λ_2 (details in Appendix E) directly by matching the first six lags of empirical and theoretical autocorrelation functions of our CARMA(2,1) process. Observe that it is equivalent to fitting a sum of two exponentials, as argued in Equation 3.9. However, this problem has been identified as ill-conditioned by many researchers, see e.g. [Kaufmann, 2003] or [Smith, Lyle B. (Computation Group, 1969]. [Acton, 1990] even pointed out that "...an exponential equation of this type [a weighted sum of two exponentials] in which all four parameters are to be fitted is extremely ill conditioned" (p. 253).

We conducted a number of numerical experiments which confirmed that in our case the estimation of the autoregressive roots is extremely difficult, if not impossible. It is probably caused by the fact that the values of roots are quite close to each other, which means that the algorithm is unable to successfully identify them, even though the fit of the full autocorrelation function is satisfactory (Figure 19). It seems that [García et al., 2011] were more lucky as the difference between their estimated roots ($\lambda_1 = -0.0465$ and $\lambda_2 = -1.9181$) is larger. We repeated our experiments using these values (in the Gaussian case) and obtained much more optimistic estimation results, with the relative bias as small as 0.056 for λ_1 and 0.067 for λ_2. Our experiments suggest that even though the algorithm of fitting CARMA(2,1) described in Appendix E is widely used in the literature, it requires more attention.

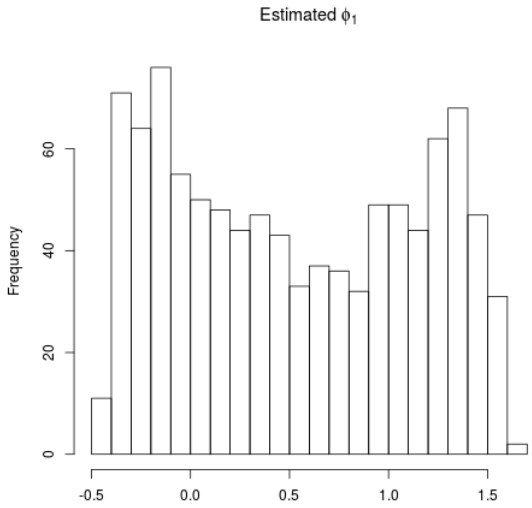

Figure 17: Estimated autoregressive parameters: ϕ_1.

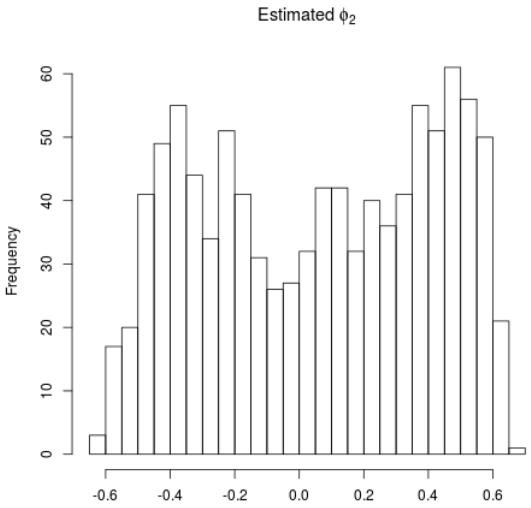

Figure 18: Estimated autoregressive parameters: ϕ_2.

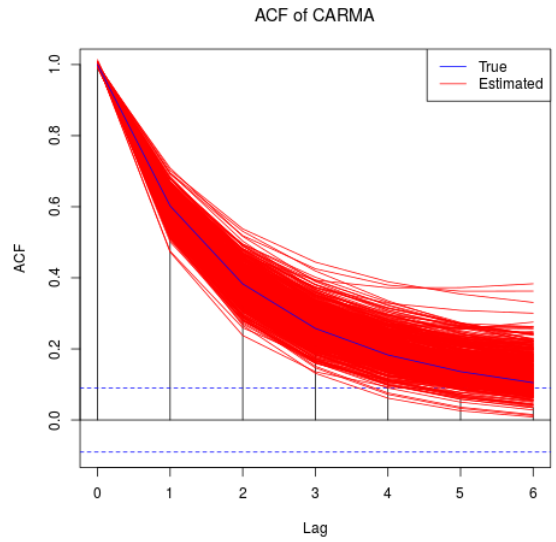

Figure 19: Autocorrelation functions of $Y(t)$: empirical (blue) and theoretical with parameters estimated from 1000 Monte Carlo simulations (red).

3.4 Stochastic volatility

As mentioned in Section 3.3.2, we should investigate if adding a stochastic volatility of $Y(t)$ would improve the model. Thus in this section again we assume that $Y(t) = \int_{-\infty}^{t} g(t-s)\sigma_s dB_s$, where $\sigma_t^2 = \int_{-\infty}^{t} e^{-\delta(t-s)} dV_s$, B_t is a standard Brownian motion, $g(\cdot)$ – the CARMA kernel and V_t – a Lévy subordinator independent of B_t.

The driving Lévy process estimated in Section 3.3.2 is now replaced by a volatility modulated Brownian motion, so $dL_t = \sigma_t dB_t$. Therefore the estimated realised variance converges uniformly on compact sets in probability to C_t, where $C_t := \int_0^t \sigma_s^2 ds$ and $C_t - C_{t-1} = \int_{t-1}^t \sigma_s^2 ds$. We compute:

- $\kappa_1 := \mathbb{E}_{\mathbb{Q}}\left[\sigma_t^2\right] = \mathbb{E}_{\mathbb{Q}}\left[V_1\right] \int_0^\infty e^{-\delta x} dx = \frac{\mathbb{E}_{\mathbb{Q}}[V_1]}{\delta}$;
- $\kappa_2 := \mathrm{Var}_{\mathbb{Q}}\left[\sigma_t^2\right] = \mathrm{Var}_{\mathbb{Q}}\left[V_1\right] \int_0^\infty e^{-2\delta x} dx = \frac{\mathrm{Var}_{\mathbb{Q}}[V_1]}{2\delta}$;
- $\mathrm{Cov}_{\mathbb{Q}}\left(\sigma_{t+h}^2, \sigma_t^2\right) = \mathrm{Var}_{\mathbb{Q}}\left[V_1\right] \int_0^\infty e^{-\delta(x+h)} e^{-\delta x} dx = \frac{\mathrm{Var}_{\mathbb{Q}}[V_1]}{2\delta} e^{-\delta h} = \kappa_2 e^{-\delta h}$.

Thus

$\bullet \mathbb{E}_{\mathbb{Q}}\left[C_t - C_{t-1}\right] = \kappa_1$;

$\bullet \mathrm{Cov}_{\mathbb{Q}}\left(C_{t+h} - C_{t+h-1}, C_t - C_{t-1}\right) = \mathbb{E}_{\mathbb{Q}}\left[\int_{t+h-1}^{t+h} \sigma_s^2 ds \int_{t-1}^{t} \sigma_u^2 du\right] - \kappa_1^2 = \int_{t+h-1}^{t+h} \int_{t-1}^{t} \mathbb{E}_{\mathbb{Q}}\left[\sigma_s^2 \sigma_u^2\right] ds du - \kappa_1^2$

$= \kappa_2 \int_{t+h-1}^{t+h} \int_{t-1}^{t} e^{-\delta|s-u|} ds du + \kappa_1^2 - \kappa_1^2 = \kappa_2 \int_{t+h-1}^{t+h} e^{-\delta u} du \int_{t-1}^{t} e^{\delta s} ds = \frac{\kappa_2}{\delta^2} \left(e^{\frac{\delta}{2}} - e^{-\frac{\delta}{2}}\right)^2 e^{-\delta h}$

$\bullet \mathrm{Var}_{\mathbb{Q}}\left[C_t - C_{t-1}\right] = \frac{\kappa_2}{\delta^2} \left(e^{\frac{\delta}{2}} - e^{-\frac{\delta}{2}}\right)^2$.

Therefore the theoretical autocorrelation function is given by $\mathrm{ACF}(h) = e^{-\delta h}$. In order to estimate the memory parameter δ, we need to match this theoretical function with its empirical counterpart, both computed for $\sum_{t=1}^{N} \left(L_t - L_{t-1}\right)^2$, where N denotes the number of observations. We decided to use the first six lags and proceed via standard linear regression, which gave us a rough estimate of $\hat{\delta} = 1.82$, at significance level 0.01. The true and estimated autocorrelation functions are presented in Figure 20. The remaining parameters were already estimated in Section 3. We conclude that including a stochastic volatility in the model for $Y(t)$ improves the model quality, as indicated by correlated squared increments of the driving process L_t as well as by the presence of a statistically significant Ornstein-Uhlenbeck memory parameter. However, in the remaining sections we will return to the model without stochastic volatility and propose a different modification so that we can identify the impact of different modifications on the model quality.

Remark 3.3. The CARMA(2,1) process is just one example of a kernel function $g(\cdot)$ fulfilling our assumptions. Following [Barndorff-Nielsen et al., 2013] one could investigate the class of gamma kernels.

If $Y(t)$ has a symmetric distribution, which can be obtained using a Box-Cox transformation (in

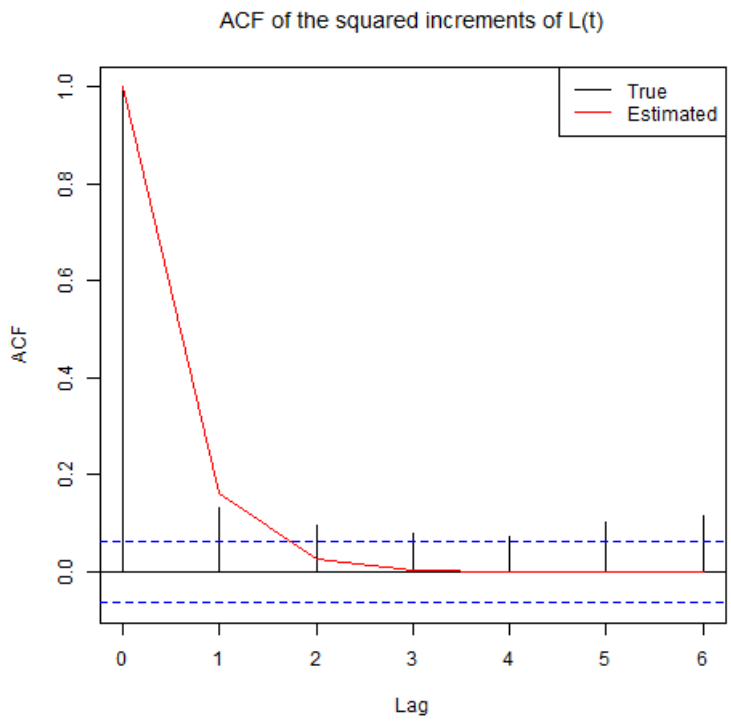

Figure 20: True and estimated ACF of the squared increments of $L(t)$.

our case with the parameter $\lambda = 1.4$), we can assume that

$$Y(t) = \mu + c \int_{-\infty}^{t} g(t-s)\sigma(s)dB(s), \tag{3.14}$$

where $c \in \mathbb{R}$ is a constant, σ is a stationary process and B denotes a standard Brownian motion independent of σ. We denote the gamma density with parameters $\bar{\nu} > 0$ and $\bar{\lambda} > 0$ by

$$\bar{g}\left(t; \bar{\nu}, \bar{\lambda}\right) = \frac{\bar{\lambda}^{\bar{\nu}}}{\Gamma(\bar{\nu})} t^{\bar{\nu}-1} e^{-\bar{\lambda}t}. \tag{3.15}$$

Let us assume that $\frac{1}{2} < \bar{\nu} < 1$. We define the gamma kernel

$$g\left(t\right) = \frac{\bar{\lambda}^{\bar{\nu}-\frac{1}{2}}}{\sqrt{\Gamma(2\bar{\nu}-1)}} t^{\bar{\nu}-1} \exp\left(-\frac{\bar{\lambda}}{2}t\right). \tag{3.16}$$

We also assume that for some subordinator $z(t)$

$$\sigma^2(t) = \int_{-\infty}^{t} i^{\star}(t-s)dz(s) \tag{3.17}$$

with

$$i^{\star}(t) = \frac{1}{\bar{\lambda}} \bar{g}\left(t; 2-2\bar{\nu}, \bar{\lambda}\right) \tag{3.18}$$

Simple calculations show that

$$Y(t)|\sigma \sim N\left(\mu, c^2\omega^2(t)\right), \tag{3.19}$$

where

$$\omega^2(t) = \int_{-\infty}^{t} e^{-\bar{\lambda}(t-s)} dz(s). \tag{3.20}$$

[Barndorff-Nielsen et al., 2013] argue that if $\omega^2(t)$ is a generalised inverse Gaussian process $GIG(\lambda, \chi, \psi)$, then $Y(t)$ has a generalised hyperbolic distribution $GH(\lambda, \chi, \psi, \mu, c^2, 0)$. [Barndorff-Nielsen et al., 2013] provide an explicit formula for the autocorrelation function of $Y(t)$

$$Cor\left(Y(t), Y(t+h)\right) = \frac{1}{\Gamma(\bar{\nu} - \frac{1}{2})} 2^{2-2\bar{\nu}} (h\bar{\lambda})^{\bar{\nu}-\frac{1}{2}} K_{\bar{\nu}-\frac{1}{2}}\left(\frac{h\bar{\lambda}}{2}\right), \tag{3.21}$$

where $K_{\bar{\nu}}(x)$ denotes the modified Bessel function of the third kind. Therefore we can estimate parameters $\bar{\nu}$ and $\bar{\lambda}$ by matching the first lags of the empirical and theoretical autocorrelation functions of $Y(t)$. The least square estimation using the first six lags resulted in the estimated $\hat{\bar{\nu}} = 0.889$ and $\bar{\lambda} = 0.845$. The empirical and estimated autocorrelation functions are presented in Figure 21. We note that for the first lags the match is very accurate, therefore this approach might be worth further investigation (for more details see for example [Bennedsen et al., 2017]).

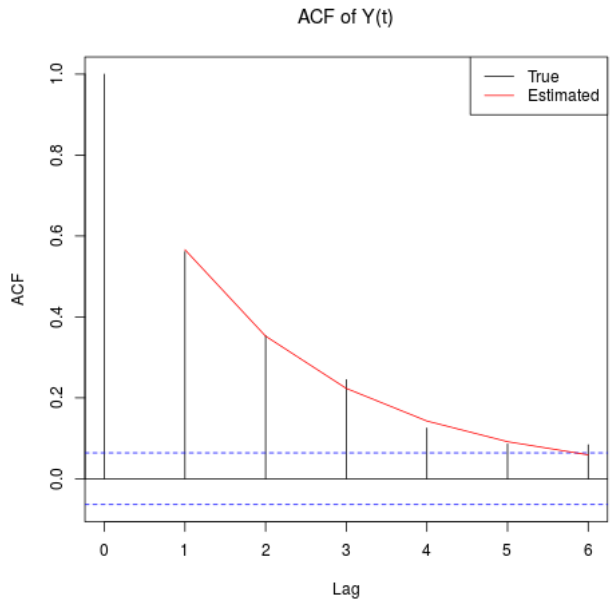

Figure 21: The empirical and estimated (based on the first six lags) autocorrelation functions of $Y(t)$ with the gamma kernel.

4 Influence of wind energy production on electricity prices

4.1 Data description

Our set of wind production data corresponds to the spot and futures prices data, i.e. we used the data from five energy providers (four from Germany and one from Austria) between 1 January 2013

and 3 August 2015. The data were obtained from the websites: German Netztransparentz [3] and Austrian Power Grid [4]. We were interested only in the global production, so we aggregated energy generated on- and off-shore as well as by all five providers. We also needed the global volumes of produced electricity, available on the EPEX SPOT website. All data were aggregated to daily values. In Table 6 we present summary statistics of the global daily wind energy production.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10816.30	67597.79	123276.52	162900.17	214119.19	766291.29

Table 6: Summary statistics of global daily wind energy generation (in MW).

4.2 Wind penetration index (WPI)

Inspired by [Veraart, 2016], we looked at the wind penetration index (WPI). We define the WPI on day t as the ratio between the day-ahead forecasted wind energy generation (WG_t) and the actual volume (V_t), i.e. $WPI_t = \frac{WG_t}{V_t}$. Since we have access only to the actual wind energy generation, we simulated its forecasts as

$$WG_t = AWG_t + \varepsilon_t, \tag{4.1}$$

where AWG_t denotes actual wind energy generation on day t and $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$ with σ_ε equal 2% of the mean volume, as suggested by [Jónsson et al., 2013]. Adding a Gaussian perturbation with a relatively small standard deviation does not change the analysis much, which justifies working with simulated rather than empirical data.

In Table 7 we present summary statistics of the wind penetration index, while its time series plot and histogram are shown in Figure 22 and Figure 23, respectively. The plot indicates that the WPI behaves seasonally, in particular it is higher during windy winters than less windy summers. From the histogram we can see that WPI is relatively low, in particular the wind energy very rarely exceeds half of the total generation.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02	0.10	0.18	0.22	0.29	0.77

Table 7: Summary statistics of the wind penetration index.

4.3 Relationship between wind energy production and spot prices

Recall that our main research goal was to determine whether (and in what way) wind energy production influences electricity prices. Figure 24 and Figure 25 present plots of the (deseasonalised)

[3] `https://www.netztransparenz.de/Erneuerbare-Energien-Gesetz/Marktpraemie` (accessed 10.11.2017)

[4] `https://www.apg.at/en/market/Markttransparenz/generation` (accessed 10.11.2017)

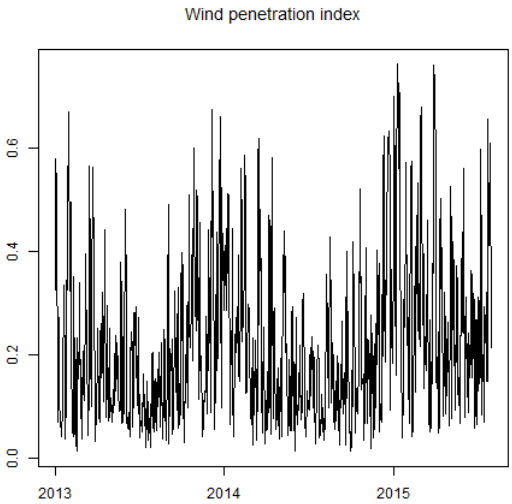

Figure 22: Wind penetration index.

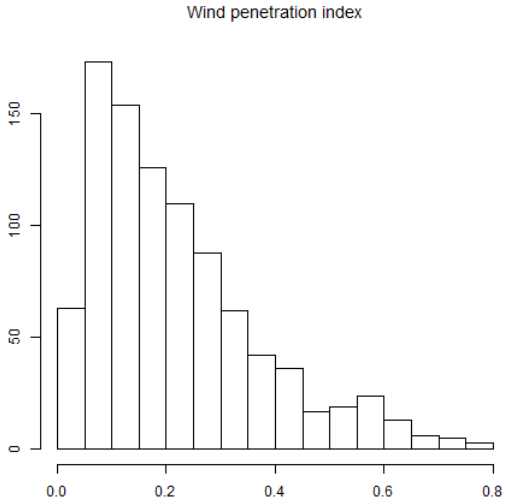

Figure 23: Histogram of the wind penetration index.

spot prices against WPI with fitted linear regression lines. There is a clear negative correlation in both cases, which means that lower spot prices (both raw and deseasonalised) correspond to high share of wind energy in total generation. The correlation coefficients equal -0.51 and -0.61 for raw and deseasonalised prices, respectively. This analysis does not tell us anything about the causation. However, since wind energy is very cheap to produce, we can expect that a high wind penetration index lowers electricity prices.

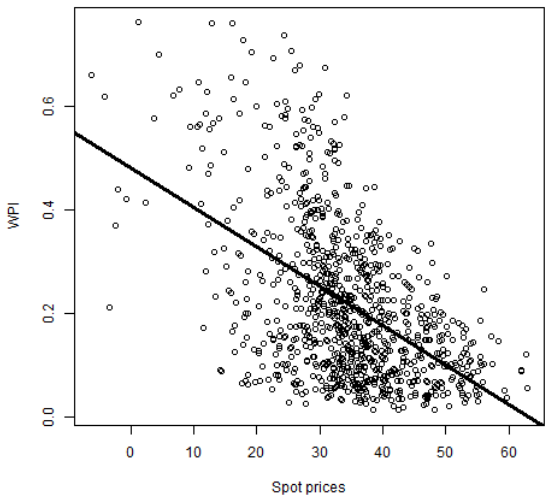

Figure 24: Spot prices vs. WPI.

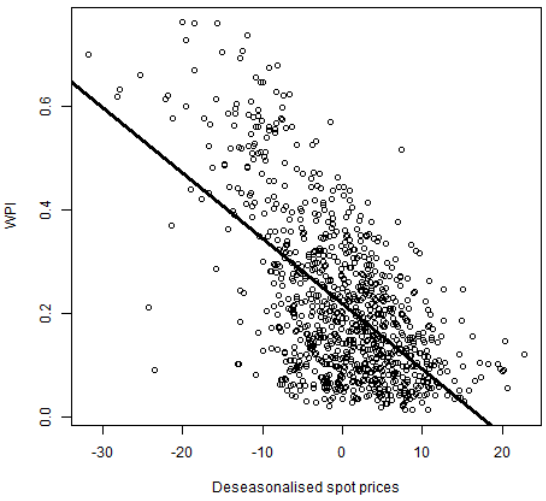

Figure 25: Deseasonalised spot prices vs. WPI.

4.4 Modified models

In this section we describe modifications of the model presented in Subsection 2.3 (without stochastic volatility) that take into account the relationship between the wind energy production and spot prices. Because long-term weather forecasts are unreliable, we assume that wind energy production influences only the short-term factor, i.e. $Y(t)$. Therefore $\Lambda(t)$ and $Z(t)$ introduced in Subsection 2.3 remain unchanged.

It is not obvious how to represent wind energy in the model. We looked at three variables: wind energy forecast F_t, wind penetration index WPI_t defined in Subsection 4.2 and the residual demand RD_t defined as the difference between the total load and the forecasted wind energy production (motivated by [Elberg and Hagspiel, 2015]). We fitted linear and quadratic models involving different linear combinations of these variables. The residuals were modelled by a CARMA(2,1) process. A similar approach was proposed by [Elberg and Hagspiel, 2015]; however, instead of parametric regression models they used spline fits, which makes simulating from the model difficult.

4.5 Comparison of models

Measuring the model performance is a difficult task. We decided to compare the model without wind energy production data (Section 3) with different variations of the model including renewables data. We defined the distance between "true" (so resulting from the split described in Section 3) and simulated processes $Y(t)$ as a sum of squared differences between first "true" moments and moments averaged from 1000 Monte Carlo simulations performed using the R package **yuima** (*distance 1*). We used different numbers of moments, from 3 to 20. Note that this definition closely resembles squared *maximum mean discrepancy* (see e.g. [Gretton, 2012]) with the feature map $\phi(x) = \left(x, x^2, \ldots, x^k\right)$, where k denotes the number of the highest moment, and a radial kernel. The squared differences were further normalised by the value of the appropriate "true" moment to reduce the influence of the large variability of higher moments on the total distance (*distance 2*). For each distance and each number of moments we defined the best model as the model minimizing the distance and the worst as the one maximizing the distance. The best and worst models are presented in Table 8. The last two columns indicate that including the wind energy production data improved the original model proposed in Subsection 2.3. The linear model with residual demand appears as the best model in Table 8 more often than any other, thus we suggest modelling the short-term process as $Y_t = a + b \cdot \mathrm{RD}_t + \mathrm{CARMA}_t$.

	Best models		Worst models	
Highest moment	Distance 1	Distance 2	Distance 1	Distance 2
3	1	2	7	5
4	5	4	8	7
5	2	5	8	8
6	5	5	1	8
7	5	5	1	8
8	5	5	1	1
9	5	5	1	1
10	5	5	1	1
11	5	5	1	1
12	9	9	1	1
13	5	5	1	1
14	5	5	1	1
15	9	6	1	1
16	5	5	7	7
17	5	5	1	1
18	9	9	1	1
19	5	5	1	1
20	6	6	1	1

Table 8: Best and worst models: models minimising and maximising distances between empirical and simulated data, for each number of moments used to calculate the metric. Models are described in Table 9.

	WPI_t	RD_t	F_t	WPI_t^2	RD_t^2	F_t^2	intercept
1	0	0	0	0	0	0	0
2	-1.828e+00	-7.559e-06	**-3.544e-05**	0	0	0	*1.032e+01*
3	0	-6.885e-06	**-3.714e-05**	0	0	0	**9.825e+00**
4	**-28.3714**	0	0	0	0	0	**6.2516**
5	0	**3.989e-05**	0	0	0	0	**-2.163e+01**
6	0	0	**-3.334e-05**	0	0	0	**5.468e+00**
7	**-14.1499**	0	0	**-23.1762**	0	0	**4.7982**
8	0	**1.538e-04**	0	0	**-1.127e-10**	0	**-4.938e+01**
9	0	0	**-3.147e-05**	0	0	-3.422e-12	**5.316e+00**

Table 9: The description of linear parts of all nine models $Y_t = a_1 + a_2 \cdot \mathrm{WPI}_t + a_3 \cdot \mathrm{RD}_t + a_4 \cdot \mathrm{F}_t + a_5 \cdot \mathrm{WPI}_t^2 + a_6 \cdot \mathrm{RD}_t^2 + a_7 \cdot \mathrm{F}_t^2$, where RD_t, WPI_t and F_t denote the residual demand, the wind penetration index and the forecasted wind production, respectively. Coefficients in bold were estimated at a significance level at least 0.001, while those in italics at a significance level at least 0.01.

5 Conclusions

In this paper we proposed a three-factor arithmetic model for electricity spot prices inspired by [Benth et al., 2014] and on this basis derived a formula for futures prices. It consists of: (a) a deterministic function representing a trend, weekly and monthly seasonalities and holiday effects, (b) a long-term factor modelled by a generalised hyperbolic Lévy process and (c) a short-term factor represented by a Lévy semistationary process. Empirical studies of German and Austrian data provided an evidence that this model is able to capture the complex behaviour of electricity prices. We generalised the model described by [Benth et al., 2014] by including a stochastic volatility in the short-term factor, as the data suggested its presence. We contributed to the field mainly by including the influence of wind energy production on the studied prices. Simulations indicated that this step significantly improved the arithmetic model which did not consider the impact of renewables. This approach seems to be very promising and further studies should be conducted. In particular, the question of how exactly to include the wind energy production data in the model is still open, even though we provided some recommendations. Furthermore, one could model the wind energy generation directly instead of treating it as an exogenous variable. Also, it would be beneficial to study other types of renewables, such as solar or hydro power, together with wind energy and traditional energy sources. Additionally one could propose different kernels of the Lévy semistationary process defining the short-term factor. Finally, the algorithm of fitting

CARMA(2,1) models should be investigated in more detail, as it definitely leaves some room for improvement.

Acknowledgements

We would like to thank Olivier Ferón for the valuable comments and insights. PR gratefully acknowledges the financial support from EPSRC under the grant EP/L016613/1 as well as from EDF. We also thank the developers of the R package **yuima** for technical support.

References

[Acton, 1990] Acton, F. S. (1990). *Numerical methods that work*. The Mathematical Association of America.

[Barndorff-Nielsen et al., 2011] Barndorff-Nielsen, O. E., Benth, F. E., and Veraart, A. (2011). Modelling electricity forward markets by ambit fields. *SSRN Electronic Journal*, pages 1–45.

[Barndorff-Nielsen et al., 2013] Barndorff-Nielsen, O. E., Benth, F. E., and Veraart, A. E. D. (2013). Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. *Bernoulli*, 19(3):803–845.

[Basse-O'Connor et al., 2014] Basse-O'Connor, A., Graversen, S., and Pedersen, J. (2014). Stochastic integration on the real line. *Theory of Probability and Its Applications*, 58(2):193–215.

[Bennedsen et al., 2017] Bennedsen, M., Lunde, A., and Pakkanen, M. S. (2017). Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 21(4):931–965.

[Benth et al., 2014] Benth, F. E., Klüppelberg, C., Müller, G., and Vos, L. (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energy Economics*, 44:392–406.

[Benth and Paraschiv, 2016] Benth, F. E. and Paraschiv, F. (2016). A space-time random field model for electricity forward prices. *Journal of Banking and Finance*, 0:1–14.

[Borovkova and Geman, 2006] Borovkova, S. and Geman, H. (2006). Analysis and modelling of electricity futures prices. *Studies in Nonlinear Dynamics & Econometrics*, 10(3):239–263.

[Breymann, 2011] Breymann, W. (2011). ghyp: A package on generalized hyperbolic distributions. *Analysis*.

[Brockwell et al., 2011] Brockwell, P. J., Davis, R. a., and Yang, Y. (2011). Estimation for non-negative Lévy-driven CARMA processes. *Journal of Business & Economic Statistics*, 29(2):250–259.

[Carmona et al., 2013] Carmona, R., Coulon, M., and Schwarz, D. (2013). Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, 7(2):167–202.

[Cartea et al., 2009] Cartea, Á., Figueroa, M. G., and Geman, H. (2009). Modelling electricity prices with forward looking capacity constraints. *Applied Mathematical Finance*, 16(2):103–122.

[Deschatre and Veraart, 2017] Deschatre, T. and Veraart, A. (2017). A joint model for electricity spot prices and wind penetration with dependence in the extremes. *SSRN: https://ssrn.com/abstract=3029318*.

[Elberg and Hagspiel, 2015] Elberg, C. and Hagspiel, S. (2015). Spatial dependencies of wind power and interrelations with spot price dynamics. *European Journal of Operational Research*, 241(1):260–272.

[García et al., 2011] García, I., Klüppelberg, C., and Müller, G. (2011). Estimation of stable CARMA models with an application to electricity spot prices. *Statistical Modelling*, 11(5):447–470.

[Gretton, 2012] Gretton, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773.

[Haar, 2010] Haar, R. (2010). *On modelling the electricity futures curve*. PhD thesis.

[Jónsson et al., 2013] Jónsson, T., Pinson, P., Nielsen, H. A., Madsen, H., and Nielsen, T. S. (2013). Forecasting electricity spot prices accounting for wind power predictions. *IEEE Transactions on Sustainable Energy*, 4(1):210–218.

[Kaufmann, 2003] Kaufmann, B. (2003). Fitting a sum of exponentials to numerical data. (x):1–9.

[Ketterer, 2014] Ketterer, J. C. (2014). The impact of wind power generation on the electricity price in Germany. *Energy Economics*, 44:270–280.

[Smith, Lyle B. (Computation Group, 1969] Smith, Lyle B. (Computation Group, S. L. A. C. (1969). PEEL - a Program to Perform Exponential "Peeling" (Fitting) On-Line. Technical report.

[Veraart, 2016] Veraart, A. (2016). Modelling the impact of wind power production on electricity prices by regime-switching Lévy semistationary processes. In Benth, F. E. and Di Nunno, G., editors, *Stochastics of Environmental and Financial Economics*, Stochastics of Environmental and Financial Economics, pages 321–340. Springer International Publishing.

A Integrability conditions

We defined the short-term process as a stochastic integral $Y(t) = \int_{-\infty}^{t} g(t-s)\sigma_{s-}dL_s$, so we need to state some additional conditions under which this integral exists. We base our discussion on the work by [Barndorff-Nielsen et al., 2013] and [Basse-O'Connor et al., 2014].

We denote the Lévy triplet of $L(t)$ associated with a truncation function $h(z) = \mathbb{1}_{\{|z|\leq 1\}}$ by (d, b, l_L), where d denotes the drift, b the variance of the Gaussian component and l_L the Lévy measure. Let us define $\phi_t(s) := g(t-s)\sigma_{s-}$. Then the process $(\phi_t(s))_{s\leq t}$ is integrable with respect to L if and only if $(\phi_t(s))_{s\leq t}$ is $\mathcal{F}$-predictable and these three conditions hold almost surely:

$$b\int_{-\infty}^{t} (\phi_t(s))^2\, ds < \infty, \tag{A.1}$$

$$\int_{-\infty}^{t}\int_{-\infty}^{\infty} \left(1 \wedge |\phi_t(s)z|^2\right) l_L(dz)ds < \infty, \tag{A.2}$$

$$\int_{-\infty}^{t} \left| d\phi_t(s) + \int_{-\infty}^{\infty} \left(h\left(z\phi_t(s)\right) - \phi_t(s)h(z)\right) l_L(dz)\right| < \infty. \tag{A.3}$$

To ensure the square integrability, we assume that $L(t)$ has a finite second moment and replace Equation A.1 by $\int_{-\infty}^{t} \mathbb{E}\left[\phi_t(s)^2\right] ds = \int_{-\infty}^{t} g(t-s)^2 \mathbb{E}\left[\sigma_s^2\right] ds < \infty$ and $\mathbb{E}\left[(g(t-s)\sigma_s ds)^2\right] < \infty$. For the latter condition is is enough to ensure that for some $a \in (0,1)$, $\int_0^\infty g^{2a}(x)dx < \infty$ and $\int_{-\infty}^{t} g^{2(1-a)}(t-s)\mathbb{E}\left[\sigma_s^2\right]$.

B Proof of Proposition 2.1

Lemma B.1. *If $\tilde{H} := \lim_{T\to\infty} H(T)$ exists in the L^2-sense, then for a fixed period $\tau > 0$ the L^2-limit $\lim_{T_1\to\infty} \frac{1}{\tau}\int_{T_1}^{T_1+\tau} H(T)dT$ also exists and equals $\tilde{H}$.*

Proof of Lemma B.1. Our assumption says that $\lim_{T\to\infty} \mathbb{E}_{\mathbb{Q}}\left[(H(T) - \tilde{H})^2\right] = 0$, so that for all $\varepsilon > 0$ there exists $\tilde{T}$ such that for all $T > \tilde{T}$ we have $\mathbb{E}_{\mathbb{Q}}\left[\left(H(T) - \tilde{H}\right)^2\right] < \varepsilon$. Thus if $T_1 > \tilde{T}$, then

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\frac{1}{\tau}\int_{T_1}^{T_1+\tau} H(T)dT - \tilde{H}\right)^2\right] \leq \max_{T_1\leq T\leq T_1+\tau} \mathbb{E}_{\mathbb{Q}}\left[\left(H(T) - \tilde{H}\right)^2\right] < \varepsilon, \tag{B.1}$$

which was to be proven. □

Proof of Proposition 2.1. First we will show that $\lim_{T_1\to\infty} \frac{1}{\tau}\int_{T_1}^{T_1+\tau}\int_{-\infty}^{t} g(T-s)\sigma_{s-}\, dL_s\, dT = 0$ in the L^2-sense. As all considered functions are measurable and non-negative, we can use Tonelli's theorem to compute

$$\lim_{T\to\infty} \mathbb{E}_{\mathbb{Q}}\left[\left(\int_{-\infty}^{t} g(T-s)\sigma_{s-}dL_s\right)^2\right] = \mathbb{E}_{\mathbb{Q}}\left[L_1^2\right]\mathbb{E}_{\mathbb{Q}}\left[\sigma_0^2\right]\int_{-\infty}^{t}\lim_{T\to\infty} g(T-s)^2 ds = 0, \tag{B.2}$$

which by Lemma B.1 proves this statement.

Now we need to prove that $\lim_{T_1\to\infty}\int_{T_1}^{T_1+\tau}\int_{t}^{T} g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right] ds\, dT = C$ in the L^2-sense. By

Lemma B.1 it is enough to prove that $\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds - C\right)^2\right] = 0$. We observe that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds - C\right)^2\right] &= \mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right)^2\right] \\ &- 2C\mathbb{E}_{\mathbb{Q}}\left[\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right] + C^2.\end{aligned} \tag{B.3}$$

Using Jensen's inequality and Tonelli's theorem, we can estimate

$$\begin{aligned}&\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right)^2\right] \geq \lim_{T\to\infty}\left(\mathbb{E}_{\mathbb{Q}}\left[\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right]\right)^2 \\ &= \lim_{T\to\infty}\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]\right]ds\right)^2 = \left(\mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]\lim_{T\to\infty}\int_0^{T-t} g(y)dy\right)^2 = C^2.\end{aligned} \tag{B.4}$$

On the other hand,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right] &= \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_{-\infty}^s e^{-\delta(s-x)}dV_x}\middle|\mathcal{F}_t\right] \leq \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_{-\infty}^t e^{-\delta(s-x)}dV_x} + \sqrt{\int_t^s e^{-\delta(s-x)}dV_x}\middle|\mathcal{F}_t\right] \\ &= \sqrt{\int_{-\infty}^t e^{-\delta(s-x)}dV_x} + \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV_x}\right] = \sigma_t e^{-\frac{\delta}{2}(s-t)} + \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV_x}\right],\end{aligned} \tag{B.5}$$

where we applied the inequality $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$ to non-negative processes as well as the identity $\int_{-\infty}^t e^{-\delta(s-x)}dV_x = e^{-\delta(s-t)}\int_{-\infty}^t e^{-\delta(t-x)}dV_x = \sigma_t^2 e^{-\delta(s-t)}$. We remark that σ_s is stationary in mean, i.e. for all $s\in\mathbb{R}$, $\mathbb{E}_{\mathbb{Q}}\left[\sigma_s\right] = \mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]$. Furthermore,

$$\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV_x}\right] = \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_0^{s-t} e^{-\delta u}dV_u}\right] \leq \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_0^{\infty} e^{-\delta u}dV_u}\right] = \mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right], \tag{B.6}$$

as we integrate a non-negative function over a smaller domain. Therefore

$$\begin{aligned}&\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]\mathbb{E}_{\mathbb{Q}}\left[\sigma_u|\mathcal{F}_t\right]\right] \leq \mathbb{E}_{\mathbb{Q}}\left[\sigma_0^2\right]e^{-\frac{\delta}{2}(s-t)}e^{-\frac{\delta}{2}(u-t)} \\ &+ \mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]\left(e^{-\frac{\delta}{2}(u-t)}\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV_x}\right] + e^{-\frac{\delta}{2}(s-t)}\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^u e^{-\delta(u-x)}dV_x}\right]\right) \\ &+ \mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^s e^{-\delta(s-x)}dV_x}\right]\mathbb{E}_{\mathbb{Q}}\left[\sqrt{\int_t^u e^{-\delta(u-x)}dV_x}\right] \leq \mathbb{E}_{\mathbb{Q}}\left[\sigma_0^2\right]e^{-\frac{\delta}{2}(s-t)}e^{-\frac{\delta}{2}(u-t)} \\ &+ \mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]^2\left(e^{-\frac{\delta}{2}(s-t)} + e^{-\frac{\delta}{2}(u-t)} + 1\right),\end{aligned} \tag{B.7}$$

where the first inequality follows from Equation B.5 and the second one from Equation B.6. This implies that

$$\begin{aligned}
&\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right)^2\right]\\
&=\lim_{T\to\infty}\int_t^T\int_t^T g(T-s)g(T-u)\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]\mathbb{E}_{\mathbb{Q}}\left[\sigma_u|\mathcal{F}_t\right]\right]dsdu\\
&\leq\mathbb{E}_{\mathbb{Q}}\left[\sigma_0^2\right]\left(\lim_{T\to\infty}\int_t^T g(T-s)e^{-\frac{\delta}{2}(s-t)}ds\right)^2+2\mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]^2\lim_{T\to\infty}\left(\int_t^T g(T-s)ds\int_t^T g(T-s)e^{-\frac{\delta}{2}(s-t)}ds\right)\\
&+\mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]^2\left(\lim_{T\to\infty}\int_t^T g(T-s)ds\right)^2=\mathbb{E}_{\mathbb{Q}}\left[\sigma_0^2\right]\left(\lim_{T\to\infty}e^{-\frac{\delta}{2}(T-t)}\int_0^{T-t}g(y)e^{\frac{\delta}{2}y}dy\right)^2\\
&+2\mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]^2\lim_{T\to\infty}\int_0^{T-t}g(y)dy\lim_{T\to\infty}\left(e^{-\frac{\delta}{2}(T-t)}\int_0^{T-t}g(y)e^{\frac{\delta}{2}y}dy\right)\\
&+\mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]^2\lim_{T\to\infty}\int_0^{T-t}g(y)dy=C^2,
\end{aligned}\tag{B.8}$$

where we used Tonelli's theorem and assumptions of Proposition 2.1. By Equation B.4 and Equation B.8 we deduce that

$$\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\left(\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right)^2\right]=C^2.\tag{B.9}$$

Because also by Tonelli's theorem

$$\lim_{T\to\infty}\mathbb{E}_{\mathbb{Q}}\left[\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]ds\right]=\mathbb{E}_{\mathbb{Q}}\left[\sigma_0\right]\lim_{T\to\infty}\int_t^T g(T-s)ds=C,\tag{B.10}$$

Equation B.3 implies that $\lim_{T_1,\to\infty}\int_{T_1}^{T_1+\tau}\int_t^T g(T-s)\mathbb{E}_{\mathbb{Q}}\left[\sigma_s|\mathcal{F}_t\right]\,ds\,dT=C$ in the L^2-sense. □

C Change of measure continued

We used the generalised Esscher transform of $L(t)$ with parameter $\theta(t)$, a Borel measurable function. One defines $\mathbb{Q}_L^\theta$ via the Radon-Nikodym density process

$$\left.\frac{d\mathbb{Q}_L^\theta}{d\mathbb{P}}\right|_{\mathcal{F}_T}=\exp\left(\int_0^t\theta(s)dL_s-\int_0^t\phi_L\left(\theta(s)\right)ds\right),\tag{C.1}$$

where $\theta(\cdot)$ is a real-valued function integrable with respect to $L(t)$ and $\phi_L(\cdot)=\log\left(\mathbb{E}[\exp(xL_1)]\right)$ is the log-moment generating function of L_1 (if it exists).

By analogy, we define $\mathbb{Q}_V^\eta$ as

$$\left.\frac{d\mathbb{Q}_V^\eta}{d\mathbb{P}}\right|_{\mathcal{F}_T}=\exp\left(\int_0^t\eta(s)dV_s-\int_0^t\phi_V\left(\eta(s)\right)ds\right);\tag{C.2}$$

Then we get a new probability measure for $Y(t)$: $\mathbb{Q}_{\mathbb{Y}} := \mathbb{Q}_L \times \mathbb{Q}_V$. We define a measure change for $Z(t)$ in a similar way, i.e.

$$\left.\frac{d\mathbb{Q}_Z^\kappa}{d\mathbb{P}}\right|_{\mathcal{F}_T} = \exp\left(\int_0^t \kappa(s)dZ_s - \int_0^t \phi_Z\left(\kappa(s)\right)ds\right). \tag{C.3}$$

Finally we define the probability measure $\mathbb{Q} := \mathbb{Q}_Y \times \mathbb{Q}_Z$. If we choose $\theta(t), \eta(t)$ and $\kappa(t)$ to be constant, then the change of measure will preserve the desirable Lévy property.

D The sampled CARMA process

Even though we assume that $Y(t)$ is a continuous process, we observe it only in discrete time. More precisely, we denote by $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$ the sampled process, where N is the number of available observations and $h > 0$ is a small, fixed interval between the consecutive observations.

Theorem D.1 ([Brockwell et al., 2011, Proposition 3])**.** *Under the assumptions of Definition 3.2 (where we replace CARMA(2,1) by a general CARMA(p,q) process) the following hold.*

1. *The sampled process $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$, with a fixed $h > 0$, can be represented as $Y_n = \sum_{r=1}^p Y_n^{(r)}$, where $n \in \mathbb{Z}$. For each $r = 1, \ldots, p$, the discrete-time process $\{Y_n^{(r)}\}$ is obtained by sampling the component CAR(1) process $\{Y^{(r)}(t)\}$ at spacing $h > 0$. As Y is strictly stationary,*

$$\forall n \in \mathbb{Z} \quad Y_n^{(r)} = e^{\lambda_r} Y_{n-1}^{(r)} + Z_n^{(r)}, \tag{D.1}$$

with the iid noise

$$\forall n \in \mathbb{Z} \quad Z_n^{(r)} = \alpha_r \int_{(n-1)h}^{nh} e^{\lambda_r(nh-s)} dL(s). \tag{D.2}$$

2. *The sampled process $\{Y_n := Y(nh), n = 0, 1, \ldots, N\}$ satisfies*

$$\phi(B)Y_n = \sum_{r=1}^p V_{n-r+1}^r =: U_n, \tag{D.3}$$

where

$$\phi(z) := \prod_{r=1}^p \left(1 - e^{\lambda_r h} z\right) = 1 - \sum_{r=1}^p \phi_r z^r \tag{D.4}$$

and B denotes the backshift operator, i.e. $B^j Y_n := Y_{n-j}$. For each $r = 1, \ldots, p$, we define the iid sequence $\{V_n^{(r)}\}$ as

$$V_n^{(r)} := \int_{(n-1)h}^{nh} \sum_{k=1}^p \alpha_k \left(e^{(r-1)h\lambda_k} - \sum_{j=1}^{r-1} \phi_j e^{(r-1-j)h\lambda_k}\right) \times e^{(nh-s)\lambda_k} dL(s). \tag{D.5}$$

3. *We can represent the right-hand side of Equation D.3 as an invertible moving average*

$$\theta(B)W_n := W_n + \theta_1 W_{n-1} + \cdots + \theta_{p-1} W_{n-p+1}, \tag{D.6}$$

where $\{W_n\}$ is a sequence of white noise (possibly not iid) and $\theta_1, \ldots, \theta_q$ are moving average constants depending on the CARMA process. Therefore $\{Y_n\}$ can be represented as a weak ARMA(p,p-1) process, so an ARMA(p,p-1) process allowing for not iid driving white noise, such that

$$\phi(B)Y_n = \theta(B)W_n \tag{D.7}$$

and

$$W_n = \theta(B)^{-1} \sum_{r=1}^{p} V_{n-r+1}^{p}. \tag{D.8}$$

Proof. For the proof we refer the reader to [Brockwell et al., 2011]. □

E Estimation of CARMA(2,1) model

We estimated CARMA parameters using the algorithm described by [García et al., 2011]. Since we are interested in CARMA(2,1), we will specify the procedure for this particular process.

1. Estimate ARMA(2,1) parameters $\boldsymbol{\beta} = (\phi_1, \phi_2, \theta)^T$, using the maximum likelihood approach. For simplicity we use the same notation for parameters and their estimators.

2. In CARMA(2,1) case Equation D.3 has the form

$$Y_n - \phi_1 Y_{n-1} - \phi_2 Y_{n-2} = \left(1 - e^{\lambda_1 h} B\right)\left(1 - e^{\lambda_2 h} B\right) Y_n. \tag{E.1}$$

   By multiplying through and matching coefficients, we obtain

$$\phi_1 = e^{\lambda_1 h} + e^{\lambda_2 h}, \quad \phi_2 = -e^{(\lambda_1 + \lambda_2)h}. \tag{E.2}$$

   This gives us a nonlinear system of two equations for the estimators of λ_1 and λ_2, whose solutions are

$$\lambda_1 = \log\left(\frac{\phi_1}{2} + \sqrt{\left(\frac{\phi_1}{2}\right)^2 + \phi_2}\right), \tag{E.3}$$

$$\lambda_2 = \log\left(\frac{\phi_1}{2} - \sqrt{\left(\frac{\phi_1}{2}\right)^2 + \phi_2}\right). \tag{E.4}$$

   From there we immediately calculate $a_1 = -(\lambda_1 + \lambda_2)$ and $a_2 = \lambda_1 \lambda_2$.

3. The right-hand side of Equation D.3 implies that the autocovariances of the process $\phi(B)Y_n$

$$\forall k \in \mathbb{Z} \quad \gamma_U(k) = \mathrm{Cov}\left(\phi(B)Y_n, \phi(B)Y_{n-k}\right). \tag{E.5}$$

   Furthermore, using Corollary 3 by [Barndorff-Nielsen et al., 2013], we calculate the autocovariance of $Y(t)$, i.e.

$$\forall k \in \mathbb{Z} \quad \gamma_Y(k) = \mathrm{Cov}\left(Y(t+k), Y(t)\right) = \int_0^\infty g(x)g(x+k)dx = w_1 e^{\lambda_1 k} + w_2 e^{\lambda_2 k}, \tag{E.6}$$

where

$$w_1 = \frac{\alpha_1^2\lambda_1\lambda_2 + {\alpha_1}^2\lambda_2^2 + 2\lambda_1\lambda_2\alpha_1\alpha_2}{2\lambda_1\lambda_2\left(\lambda_1+\lambda_2\right)}, \tag{E.7}$$

$$w_2 = \frac{\alpha_2^2\lambda_1\lambda_2 + {\alpha_2}^2\lambda_1^2 + 2\lambda_1\lambda_2\alpha_1\alpha_2}{2\lambda_1\lambda_2\left(\lambda_1+\lambda_2\right)}. \tag{E.8}$$

Thus the autocorrelation of $Y(t)$ equals

$$\forall k \in \mathbb{Z} \quad \delta_Y(k) = \frac{\gamma_Y(k)}{\gamma_Y(0)} = \frac{w_1 e^{\lambda_1 k} + w_2 e^{\lambda_2 k}}{w_1 + w_2}. \tag{E.9}$$

For CARMA(2,1) Equation E.5 can be written as

$$\gamma_U(0) = (1+\phi_1^2+\phi_2^2)\gamma_Y(0) + (2\phi_2\phi_2 - 2\phi_1)\gamma_Y(1) - 2\phi_2\gamma_Y(2), \tag{E.10}$$

$$\gamma_U(1) = -\phi_2\gamma_Y(3) + \phi_1(\phi_2-1)\gamma_Y(2) + (1+\phi_1^2+\phi_2^2-\phi_2)\gamma_Y(1) + \phi_1(\phi_2-1)\gamma_Y(0), \tag{E.11}$$

where we can use explicit formulae for $\gamma_Y(\cdot)$ given by Equation E.6. Since they depend on a_1, a_2 and b_0, we plug in the estimates of the first two parameters.
On the other hand, the autocorrelation function at lag 1 of a moving average process with coefficient θ can be expressed as

$$\delta_U(1) = \frac{\gamma_U(1)}{\gamma_U(0)} = \frac{\theta}{1+\theta^2}. \tag{E.12}$$

Now we can replace the left-hand side of Equation E.12 by expressions from Equation E.10 to get a non-linear equation for b_0, which we solve numerically.

4. Having estimated the parameters of CARMA(2,1), we need to recover the background driving Lévy process L(t). We will use results from Section 5 of [Brockwell et al., 2011].

$$X^{(0)}(t) = X^{(0)}(0)e^{-b_0 t} + \int_0^t e^{-b0(t-s)}Y(s)ds, \tag{E.13}$$

$$X^{(1)}(t) = DX^{(0)}(t) = -b_0 X^{(0)}(t) + Y(t). \tag{E.14}$$

The canonical state vector $\mathbf{Y}(t)$ is given by

$$\begin{pmatrix} Y^{(1)}(t) \\ Y^{(2)}(t) \end{pmatrix} = \frac{1}{\lambda_1-\lambda_2} \begin{pmatrix} -\lambda_2(b_0+\lambda_1) & (b_0+\lambda_1) \\ \lambda_1(b_0+\lambda_2) & -b_0+\lambda_2 \end{pmatrix} \begin{pmatrix} X^{(0)}(t) \\ X^{(1)}(t) \end{pmatrix}. \tag{E.15}$$

To recover the background driving Lévy process L(t) we can choose one of two equations, either with $r=1$ or $r=2$:

$$L(t) = \frac{1}{\alpha_r}\left[Y^{(r)}(t) - Y^{(r)}(0) - \lambda_r \int_0^t Y^{(r)}(s)ds\right]. \tag{E.16}$$

The recommended choice is r such that $|\lambda_r|$ is minimal, which minimizes the contribution of $\lambda_r \int_0^t Y^{(r)}(s)ds$ compared to $Y^{(r)}(t) - Y^{(r)}(0)$, as discussed by [Brockwell et al., 2011, Example 4].