

Stochastic Modelling of Photovoltaic Power Generation and Electricity Prices

Fred Espen Benth^a and Noor 'Adilah Ibrahim^{b,c}

^a*Department of Mathematics, University of Oslo, PO Box 1053 Blindern, N-0316 Oslo, Norway*

^b*Department of Mathematics, University of Oslo, PO Box 1053 Blindern, N-0316 Oslo, Norway*

^c*Faculty of Science and Technology, Islamic Science University of Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia*

March 6, 2017

Abstract

In recent years, renewable energy has gained importance in producing power in many markets. The aim of this article is to model photovoltaic (PV) production for three transmission operators in Germany. PV power can only be generated during sun hours and the cloud cover will determine its overall production. Therefore, we propose a model that takes into account the sun intensity as a seasonal function. We model the deseasonalized data by an autoregressive process to capture the stochastic dynamics in the data. We present two applications based on our suggested model. First, we build a relationship between electricity spot prices and PV production where the higher the volume of PV production, the lower the power prices. As a further application, we discuss virtual power plant derivatives and energy quanto options.

1 Introduction

Renewable power generation such as wind, solar and hydropower are gaining importance as an alternative to fossil fuels. Developing countries like Pakistan, as reported by Khalil & Zaidi [26], are facing energy shortfall as they are unable to import fossil fuels in sufficient amounts. European countries, like Germany, has significantly increased their renewable power production from wind, biomass and photovoltaic (PV), the so-called *Energiewende*, in order to reduce carbon emissions. There are challenges with renewable power generation such as wind and PV as these are intermittent and storage of over-production is highly inefficient. This has led to price spikes and even negative prices in some markets.

PV power generation is highly dependent on sun intensity and cloud cover, and thus fluctuates over time. Predicting the exact amount of sunlight absorption by the solar panels is a challenging task. In order to develop a model for the production of PV power, we need to take into account the randomness induced by sun intensity and cloud cover. We propose a model with two components measuring the production; a deterministic and a stochastic. The deterministic part is representing the long-term mean characteristics while the stochastic part reflects the short time random fluctuations. Many extensive simulation studies have been conducted in investigating the performance of the PV system, but as far as we know relatively few modelling the stochastic dynamics of production in financial applications. In this paper we focus our efforts on modelling the PV production in three transmission system operators in Germany, being 50Hertz, Amprion and TransnetBW, for which we have available time series data.

Veraart & Zdanowicz [28] applied a classical trigonometric seasonality function to model the mean PV production on a similar data set as ours. We improve their approach by including the sun

intensity as an explanatory factor in modelling the seasonality of PV. This choice turns out to provide a remarkable fit of the mean variation of PV production in Germany. The sun hours is the main part in generating PV power, however the cloud cover will impact the overall production. Therefore, in the first part of this paper, we propose a model of PV production as a combination of deterministic and stochastic function. Since determining the cloud cover is a strenuous task, we model it implicitly as a stochastic component being an autoregressive process. This autoregressive process will have non-Gaussian seasonally varying residuals. We propose a seasonal variance function, that explains the seasonality very well, and propose a skew normal distribution to model the standardized residuals.

There are other approaches to stochastic modelling of PV like Yao [29] who applied logit analysis and stochastic processes to determine the energy performance. Hocaoglu [24] proposed hidden Markov processes to model the solar radiation and Cococcioni *et al.* [15] considered time series analysis and neural networks in building a model of PV generation. All these models require data of irradiation and cloud motion.

As for the application of our proposed model, we introduce a joint stochastic model for the power price and PV production. Further, we introduce an application to risk management in the power market, discussing quanto options and virtual power plant contracts. We also suggest a continuous time version of our stochastic dynamics (using so-called continuous-time autoregressive processes), that may be beneficial when pricing exotic derivatives like the above.

The rest of the paper is organized as follows: In Section 2, we carry out an empirical analysis of German PV production by looking at the trend and seasonal component, an ARMA process to model the stochastic behavior and fit the residuals with an appropriate distribution. In the next section, we analyze the relationship between power spot prices and PV production. In Section 4, we discuss an application to PV based options. Finally, Section 5 concludes the paper.

2 Empirical analysis of PV production

2.1 Data description

We analyze three Transmission System Operators (TSOs) in Germany: 50Hertz, Amprion and TransnetBW, where we have accessible PV production data ranging from 1 January 2012 to 31 December 2015. The daily data measures the PV production at 12:00 to 12:15, which we call the 12pm production. The 12pm production gives a measure for the maximal PV production. We have 1460 12pm production data for each TSOs after eliminating the observation on February 29 in a leap year. All missing values (approximately 1% of total data set) are filled by using linear interpolation of the closest neighboring production values. The areas covered by 50Hertz, Amprion and TransnetBW are shown in Figure 1. The red area on the map is the TSO TenneT, which is not covered in this paper. We use the latitude and longitude for the centers of operations for the three TSOs as proxies for their locations, being Berlin for 50Hertz, Pulheim for Amprion and Wendlingen for TransnetBW. The respective latitudes and longitudes for the TSOs locations ¹ are reported in Table 1.

Table 1: Latitude and longitude for respective locations

TSOs	Latitude	Longitude
50Hertz	52.5	13.4
Amprion	51.0	6.8
TransnetBW	48.7	9.4

The correlation between each pairs of production data of the three TSOs are presented in Table 2. The data series are seen to be strongly correlated with each other. The shorter the distance, the

¹<http://www.latlong.net/>

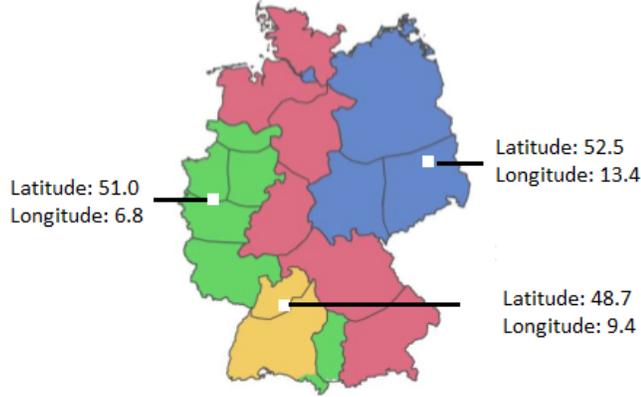


Figure 1: Transmission system operators in Germany (blue - 50Hertz, green - Amprion, yellow - TransnetBW)

higher the correlation in which the distance between Amprion and TransnetBW is the shortest (416.1 km) with 50Hertz and Amprion (502.1 km) and, 50Hertz and TransnetBW (640.5 km) following.

Table 2: Correlation between pairs of TSOs Photovoltaic Production

TSOs	50Hertz	Amprion	TransnetBW
50Hertz	-	0.806	0.725
Amprion	0.806	-	0.893
TransnetBW	0.725	0.893	-

The descriptive statistics is presented in Table 3. The average 12pm production for 50Hertz, Amprion and TransnetBW are 2445.62 MW, 2714.96 MW and 1772.33 MW, respectively. Meanwhile, we observe that the data are skewed to the right, and the negative kurtosis clearly suggests that the data are not normally distributed.

Table 3: Descriptive statistics for 50Hertz, Amprion and TransnetBW

TSOs	Min	Max	Mean	Std Dev	Skewness	Kurtosis
50Hertz	66	5866	2445.62	1490.07	0.181	-1.096
Amprion	43	6297	2714.96	1485.39	0.197	-0.940
TransnetBW	25	3842	1772.33	1029.37	0.140	-1.174

2.2 Trend and Seasonal Component

In analyzing the data, we follow a step by step procedure. First, we look at the presence of the trend and seasonality components, followed by estimating an AR process to fit the stochastic dynamics in the deseasonalized data. Then we carry out residuals analysis to detect their probabilistic properties. Figure 2 and Figure 3 show the production time series and their autocorrelation function (ACF), respectively, both displaying very strong seasonal effects. To model this, we let $PV(t)$ be PV production at time $t = 0, 1, 2, \dots, 1460$ where time is measured in days (referring to 12pm), and suppose it is a composition of a deterministic seasonal function, $\Lambda(t)$, and stochastic component, $X(t)$, given as follows,

$$\ln(PV(t)) = \Lambda(t) + X(t) \quad (2.2.1)$$

where

$$\Lambda(t) = a + bt + c \ln(I(t)). \quad (2.2.2)$$

Here, a is the average level of the PV production, b is the growth rate (due to expanded capacity), and $I(t)$ represent the sun intensity which will model the cyclical pattern with c as a scaling parameter. We will discuss a suitable model for the random part $X(t)$ in the next subsection.

Several previous studies of temperature variations used a trigonometric cyclical function with yearly periodicity to model the seasonality component (see Benth *et. al* [6], Benth & Šaltytė Benth [7], Benth & Taib [10] and Härdle & Lopez Cabrera [20]). A study by Veraart & Zdanowich [28] used trigonometric cyclical function to model PV production. Rather than a potentially high-order sum of trigonometric functions, we propose to use the sun intensity function instead which has a clear physical motivation. The sun intensity function takes into consideration the air mass and earth angle which will tell us how much of the sun radiation is absorbed and henceforth how much sun energy is absorbed by the solar panel. Before going into details with the specification of $I(t)$, we present the PV production and the fitted seasonal function Λ using the sun intensity $I(t)$ in Figure 2, which illustrates that the proposed function is suitable to model our data.

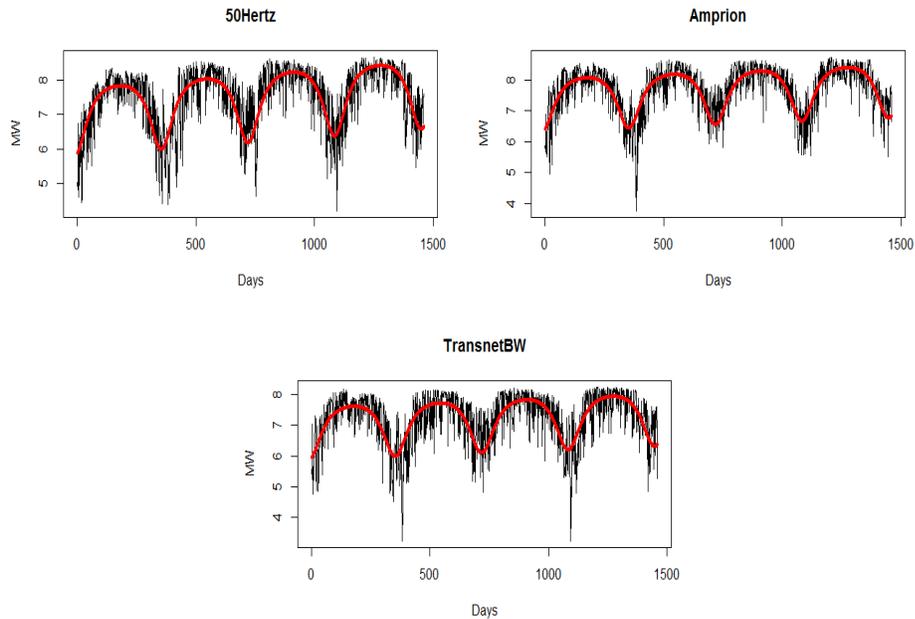


Figure 2: Time Series of 12pm production with fitted function (red - $\ln(I(t))$, black - $\ln(PV(t))$).

We present the sun intensity function $I(t)$ following Duffie & Beckman [17] and Honsberg & Bowden [25]. The sun intensity function, measured in units of kW/m^2 , is

$$I(t) = 1.353 \times 0.7^{AM^{0.678}}, \quad (2.2.3)$$

where 0.7 refers to that 70% of the sun radiation is transmitted to the earth, the figure 0.678 is the empirical fit to observed data and non-uniformities in the atmospheric layers (see Duffie & Beckman

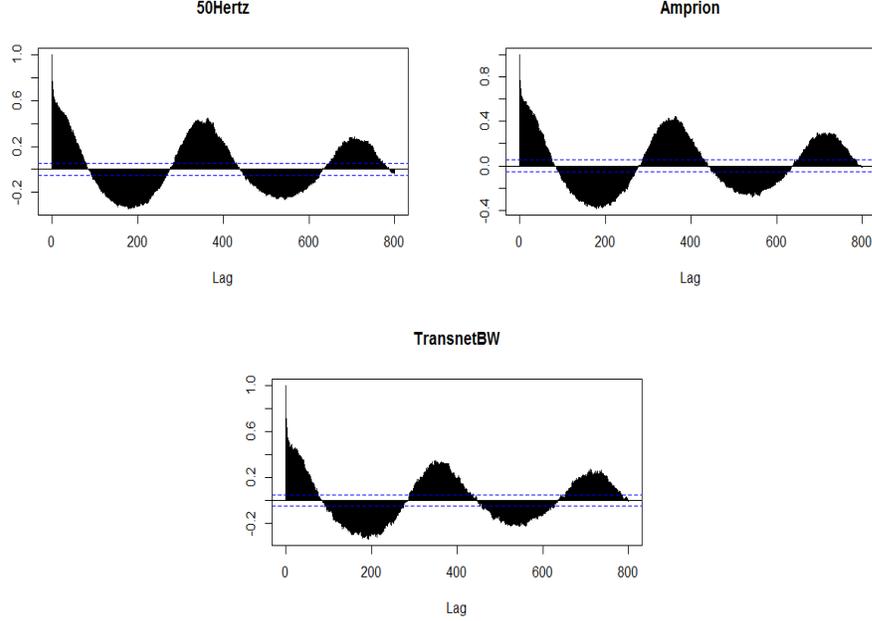


Figure 3: ACF plot

[17] and Honsberg & Bowden [25] for details), while AM denotes the air mass function. The air mass function is dependent on the sun ray angle and the curvature of the earth, taking the form

$$AM = \frac{1}{\cos(\theta) + 0.50572(96.07995 - \theta)^{-1.6364}}. \quad (2.2.4)$$

Here, θ is the zenith angle which is measured as the angle from vertical. For instance, if the sun is directly overhead, θ will be zero and the air mass is very close to 1. This indicates that the radiation is relatively little absorbed by the surroundings. The air mass function will measure the reduction of light power which passes through the atmosphere and is absorbed by air and dust. The larger the value of θ , the larger the value of the air mass. Notice that for a given location on earth (like an area of a TSO), the zenith angle, θ , will vary with time. Thus, AM is time dependent.

The zenith angle must be between 0° and 90° (when the sun is directly overhead) defined via the relationship

$$\cos \theta = \cos \phi \cos \delta \cos \omega + \sin \phi \sin \delta, \quad (2.2.5)$$

where ϕ is the latitude, δ is the declination angle, and ω is the hour angle. The parameter δ varies seasonally due to the tilting of the earth on its axis, and its value will vary between $+23.45^\circ$ and -23.45° . The parameter δ can be computed from

$$\delta = 23.45^\circ \sin\left[\frac{360}{365}(284 + n)\right],$$

with n being the n th day of the year. The next term needed to get θ is ω , which is found by:

$$\omega = 15^\circ(LST - 12\text{pm}).$$

Here, 15° refers to the angular speed of rotation of the earth, which is 15° per hour. LST is the Local Solar Time. It does not coincide with the Local Time (LT), so we need to convert our local time to the solar time by using Time Correction (TC), in minutes,

$$LST = LT + TC$$

where

$$TC = 4(LSTM - Longitude) + E.$$

$LSTM$ denotes the standard meridian of local time zone with ΔT_{GMT} being the difference of local time from Greenwich Mean Time (GMT),

$$LSTM = 15^\circ \Delta T_{GMT},$$

and E is Equation of Time, in minutes, given by

$$E = 229.2(0.000075 + 0.001868 \cos B - 0.032077 \sin B - 0.014615 \cos 2B - 0.04089 \sin 2B),$$

where

$$B = (n - 1) \frac{360}{365}.$$

We observe that GMT in Germany is 1 hour and therefore the $LSTM$ is 15° . Hence, we have the intensity function $I(t)$ in (2.2.3) explained via AM in (2.2.4) and θ in (2.2.5).

In our seasonality function (2.2.2), we model the cyclicity of logarithmic PV production by $c \ln(I(t))$, that is, proportional to the logarithmic sun intensity. Here, $I(t)$ will refer to the area of the TSO in question, and the scale c converts the “maximally accessible” sun energy into actual production. The conversion factor c takes into account the area of the solar panels, quality of the cells, locations, etc.

Using the lm function in R, we fit the linear trend and intensity function to the data, and the results are shown in Table 4. All parameter estimates are highly significant, with $p < 2e - 16$. The positivity of the estimated b parameters indicates the upward trend in the mean level of photovoltaic production for Germany. We also observe that the parameters a and c are reasonably stable across the three TSOs.

Table 4: Estimated parameters for linear trend and seasonal function.

TSOs	\hat{a}	\hat{b}	\hat{c}
50Hertz	7.660	5.42e-04	2.752
Amprion	7.934	3.04e-04	2.617
TransnetBW	7.451	2.99e-04	3.185

2.3 An autoregressive model

Our proposed deterministic function $\Lambda(t)$ captures the seasonality very well. Next we move on to model the stochastic part. We first eliminate the linear trend and seasonal component in the data. The deseasonalized PV power is obtained by subtracting the original data with the seasonality, $\Lambda(t)$, function as below:

$$X(t) = \ln(PV(t)) - \Lambda(t) \tag{2.3.1}$$

To determine the variations of $X(t)$, we plot the ACF and the partial ACF (PACF) of the deseasonalized PV data in Figure 4. Apparently, the seasonality features of the data have been removed, and we are left with an autoregressive structure.

The PACF plot in Figure 4 shows that an AR(2) process should be enough to explain the autoregressive structure in 50Hertz as well as Amprion data, while TransnetBW needs one lag higher. For consistency, we use an AR(3) process for all three TSOs. Besides that, the moving average (MA) is not significant in our case, and therefore we only take into account the autoregressive (AR) part. Thus, we assume the dynamics

$$X(t) = \beta_1 X(t - 1) + \beta_2 X(t - 2) + \beta_3 X(t - 3) + \varepsilon(t) \tag{2.3.2}$$

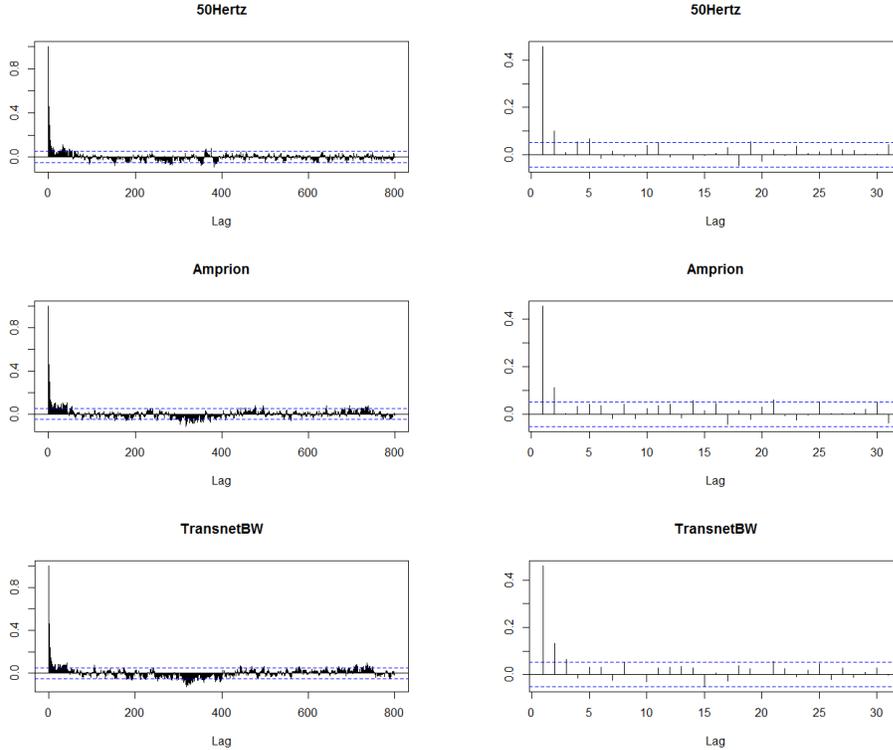


Figure 4: (Left) ACF, (Right) PACF of deseasonalized data

where $\beta_1, \beta_2, \beta_3$ are constants and $\varepsilon(t), t = 1, 2, \dots$ are independent, identically distributed (i.i.d) random variables. We refer to these as the residuals of the AR(3) model.

The estimated regression parameters of AR(3) are reported in Table 5. Standard software in R was applied to find these estimates. Since the moduli of the roots of the autoregressive polynomial are outside the unit circle in all three cases, the fitted AR(3) models are stationary. We note in passing that Veraart & Zdanowicz [28] used an ARMA(2,1) process for the deseasonalized data, where the moving average component is significant.

Table 5: Regression parameters of the AR(3) processes

TSOs	β_1	β_2	β_3
50Hertz	0.411	0.102	0.008
Amprion	0.405	0.112	0.004
TransnetBW	0.394	0.107	0.063

2.4 Residuals

Next, we proceed with the analysis of the residuals $\varepsilon(t)$. Clearly, the plots in Figure 5 give a sign of seasonal heteroskedasticity in the volatility. In addition, very small p-values for all three TSOs in Table 6 supported by the left skewed density and Q-Q plots in Figure 6 give us a strong reason to reject the null hypothesis of normality. The descriptive statistics for the residuals are reported in Table 6.

In Figure 7 we plot the ACF of the squared residuals and observe a clear seasonal pattern. Thus,

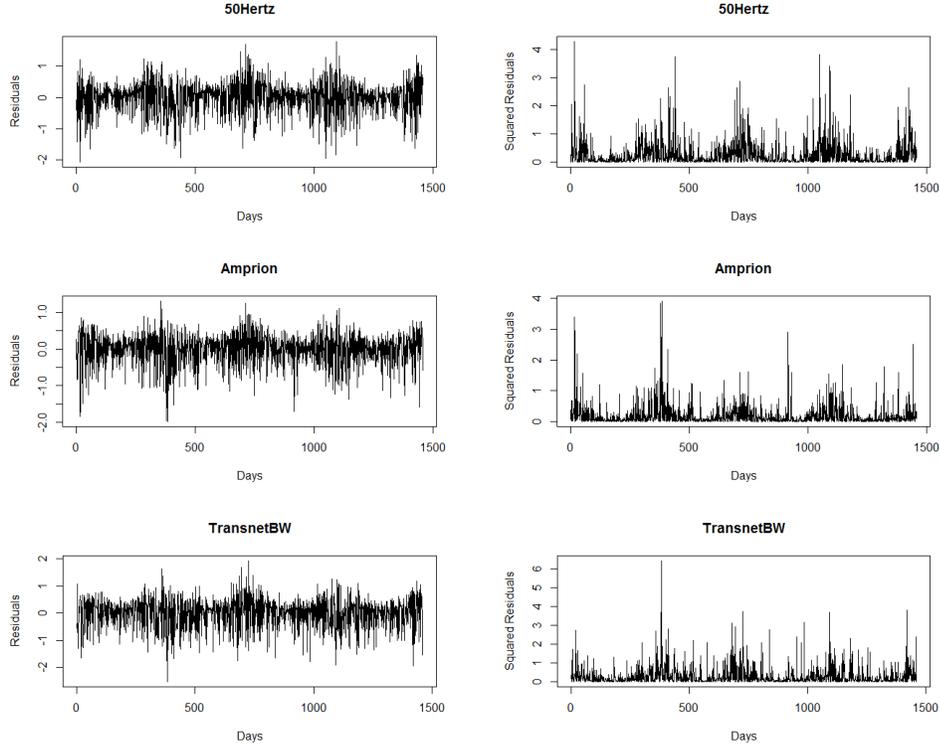


Figure 5: (Left) Residuals. (Right) Squared Residuals

Table 6: K-S values, p-values, Skewness and Kurtosis (Residuals)

TSOs	K-S statistics	p-value	Skewness	Kurtosis
50Hertz	0.074	2.519e-07	-0.533	0.930
Amprion	0.077	5.891e-08	-0.817	1.348
TransnetBW	0.099	1.003e-12	-0.665	0.899

we suggest to use a deterministic truncated Fourier series to explain the cyclical dynamics of the residual variance, motivated by the ideas of Benth & Šaltytė Benth [8] who apply this approach to model a similar phenomenon in temperature and wind speed data. Following Benth & Šaltytė Benth [8], we let,

$$\varepsilon(t) = \sigma(t)\tilde{\varepsilon}(t),$$

where $\sigma(t)$ is the square root of a truncated Fourier series,

$$\sigma^2(t) = b_1 + \sum_{l=1}^L (b_{2l}\cos(2l\pi t/365) + b_{2l+1}\sin(2l\pi t/365)), \quad (2.4.1)$$

and $\tilde{\varepsilon}(t)$ are the *standardized* residuals with $t = 1, 2, \dots, 1460$.

To fit the function $\sigma^2(t)$, we collect the squared residuals of the first week in each year, average it and label as Group 1 (28 data per week; 2012 until 2015). We continue doing this for each week throughout a year until we reach the last week; week 52, and label it as Group 52. Altogether we will have 52 averaged squared residual data and the purpose behind the averaging into weekly is to get nicer data for estimating the volatility. We then follow the approach in Benth & Šaltytė Benth

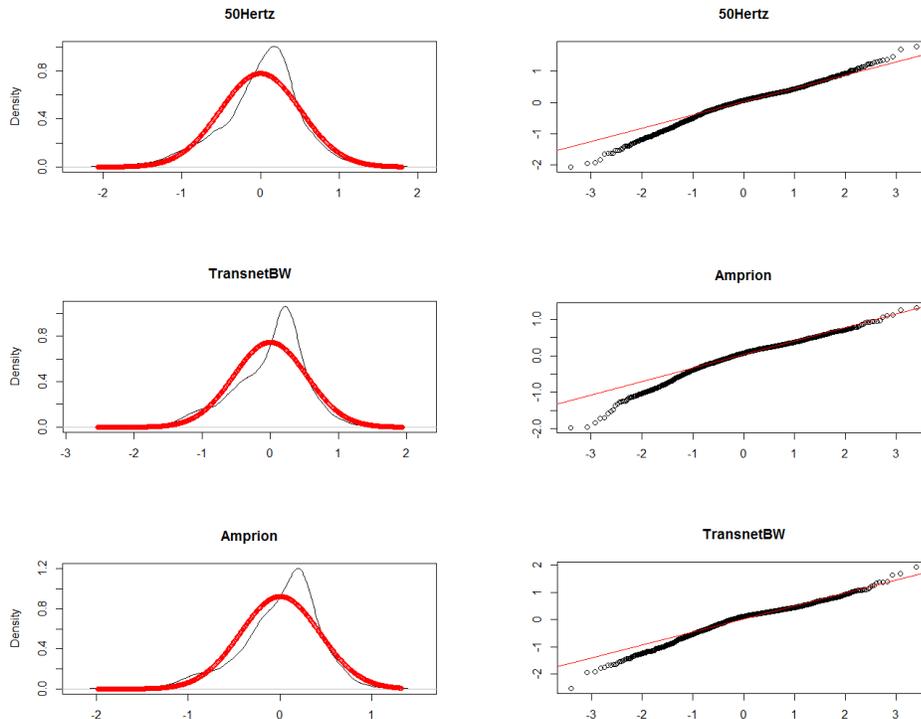


Figure 6: Density of Residuals

[8] to fit $\sigma^2(t)$ to our data by using nonlinear least squares approach, that is, using the *nonlinear least square* function in R. Notice that since we are averaging over a week, we use the middle of the week as time samples in the fitting of the function $\sigma^2(t)$, that is, we use t equal to Thursday in the week in question. We choose $L = 1$ in the specification of $\sigma^2(t)$ in (2.4.1), since the p-values for the estimated parameters of $L > 1$ are insignificant and do not give much improvement. The fitted parameters and plots of the average squared residuals are displayed in Table 7 and Figure 8, respectively. The results show that the variations during winter is higher than summer. We compute the standardized residuals, $\tilde{\varepsilon}(t)$, by dividing each residual with the volatility $\sigma(t)$ for the given day t .

The ACF plots of the standardized residuals are presented in Figure 10, which show a variation around zero indicating zero correlated data. The ACF of the squared standardized residuals is plotted in Figure 11, where we see clearly that the seasonality observed in Figure 7 has been explained by the volatility function $\sigma(t)$.

Table 7: Fitted Parameters of $\sigma^2(t)$.

TSOs	b_1	b_2	b_3
50Hertz	5.451	-4.841	-2.319
Amprion	3.522	-3.049	-1.615
TransnetBW	4.683	-4.085	-2.002

The density of the standardized residuals are depicted in Figure 9, which are clearly still non-normal. The p-values found in Table 8 are very significant, and we reject the null hypothesis of normality. Rather than the normal distribution, we suggest to use the skewed normal distribution

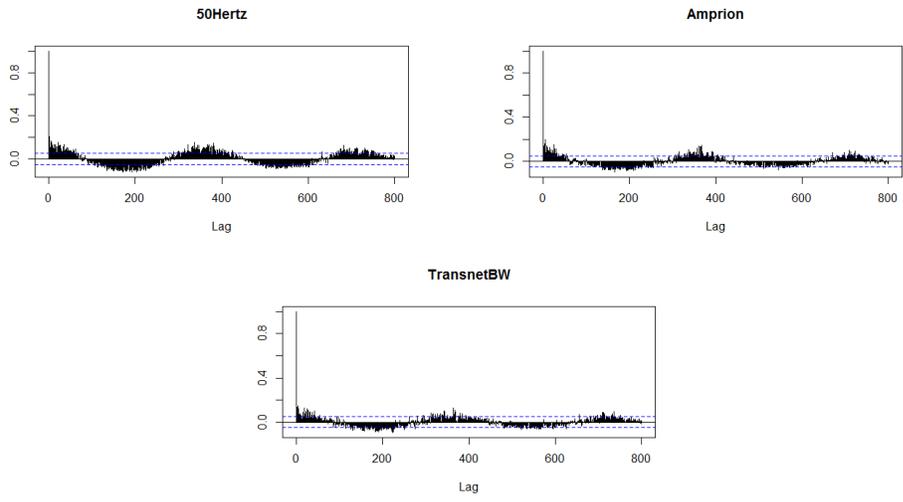


Figure 7: ACF of Squared Residuals

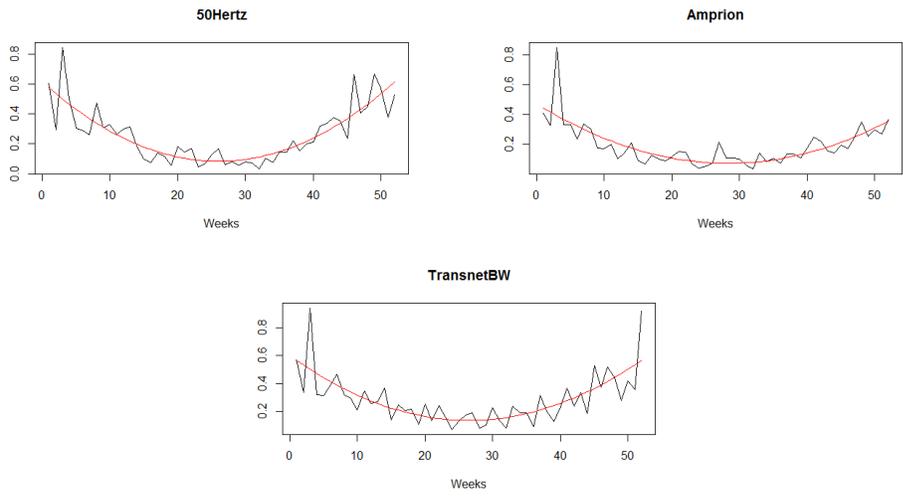


Figure 8: Fitted Average Squared Residuals

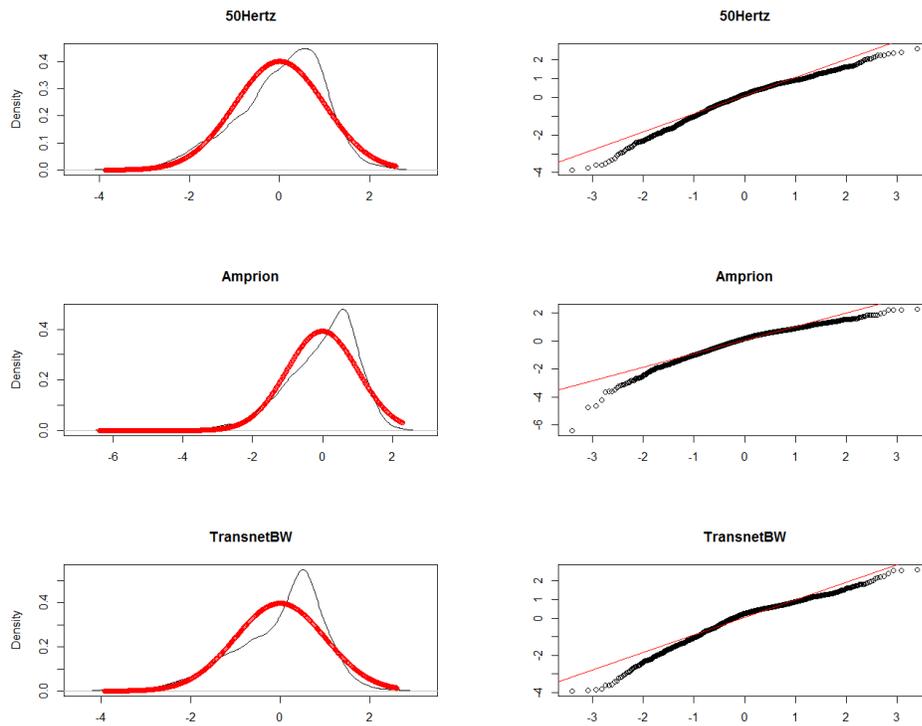


Figure 9: (Left) Density plot. (Right) Q-Qplot of Standardized Residuals

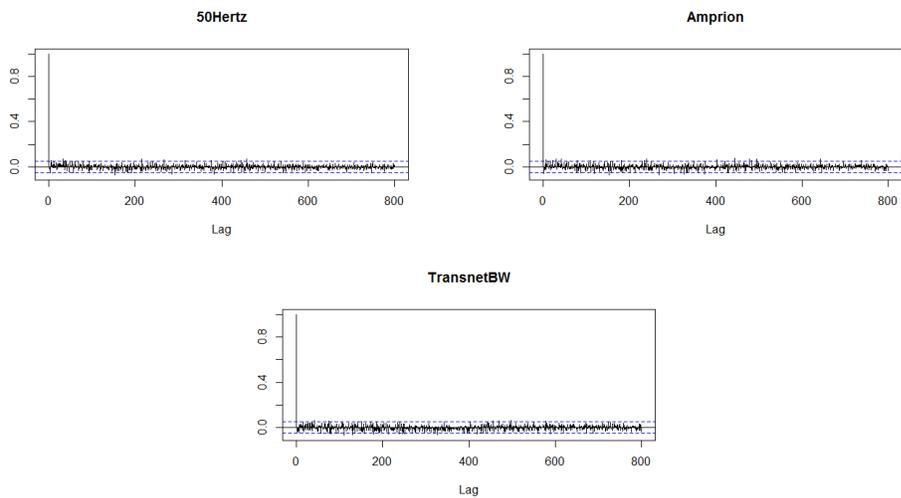
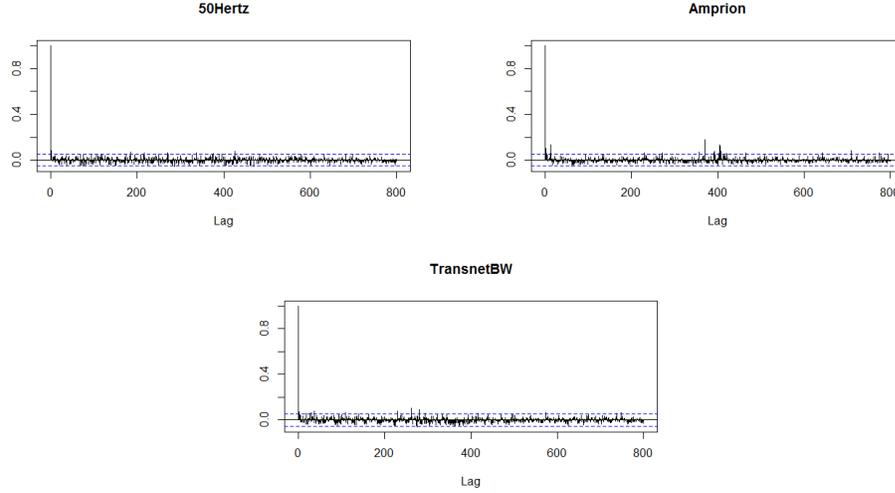


Figure 10: ACF of Standardized Residuals

Table 8: Descriptive Statistics of Standardized Residuals

TSOs	K-S statistics	p-value	Skewness	Kurtosis
50Hertz	0.071	8.194e-07	-0.690	0.511
Amprion	0.080	1.261e-08	-0.995	1.912
TransnetBW	0.105	1.898e-14	-0.827	0.789

**Figure 11:** ACF of Squared Standardized Residuals

to fit the standardized residuals.

This distribution was introduced by Azzalini [2] and categorized in the skew-elliptical class which may preserve the special characteristics of the normal distribution but also adding new benefits in regulating the skewness. Some studies in finance and insurance have applied this distribution (see Eling [18] and Bolancé [13]). Its probability density function has three parameters ξ , ω , and α , and is explicitly given by

$$f(x; \xi, \omega, \alpha) = 2\phi(x)\Phi(\alpha x), \quad (2.4.2)$$

where ξ , ω and α are location (mean), scale (standard deviation, which should be greater than 0) and shape parameters, respectively, whereas $\phi(\cdot)$ denotes the standard normal density and $\Phi(\cdot)$ its distribution function. We fit the skewed normal distribution to our standardized residuals data by maximum likelihood estimation (MLE). We applied the *dsm* function in R for this purpose.

Table 9: Estimated parameters of the skewed normal distribution

TSOs	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$
50Hertz	1.139	1.509	-2.946
Amprion	1.209	1.575	-3.778
TransnetBW	1.152	1.528	-3.250

In Table 9, we show the estimated parameters of skewed normal distribution. The standardized residuals for all three TSOs are located from 1.14 to 1.21. In Figure 12, we plot the fitted skewed normal distribution together with the empirical density demonstrating a rather decent fit.

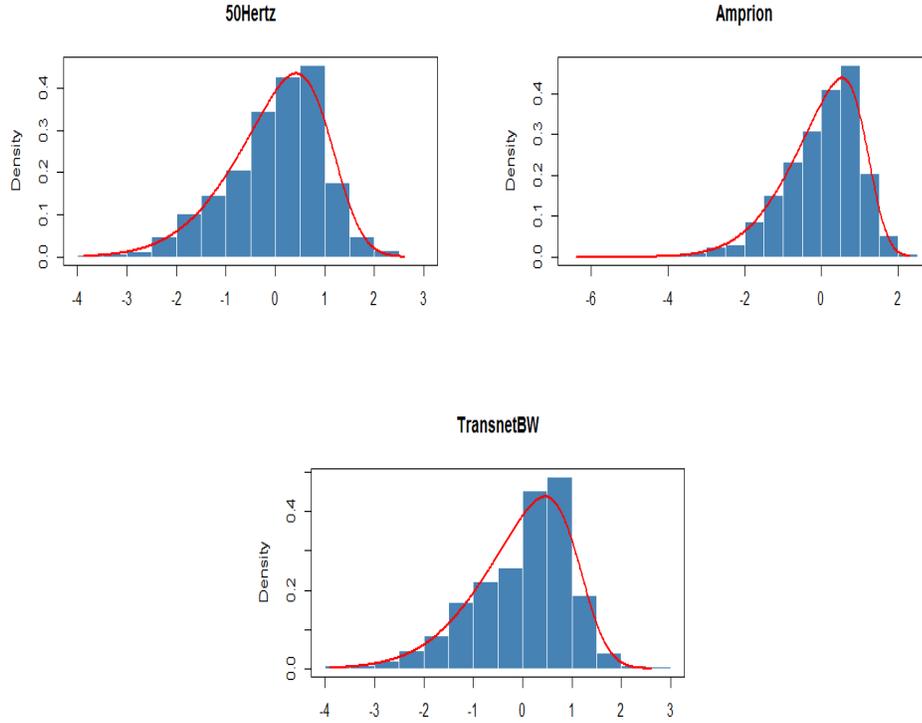


Figure 12: Fitted skewed normal distribution

3 Electricity Spot Prices Modelling

In this section, we analyse German power spot prices based on day-ahead base load between 12pm to 1pm prices quoted at the EPEX from 2012 until 2015. In particular, our goal is to get a relationship between power prices and PV production. PV is an important factor driving electricity spot prices. Generally speaking, the higher the PV production, the lower the prices, and thus PV production also impacts power generation from traditional fuels like coal, gas and oil.

In Figure 13 one can see the data series of power prices that we have available for analysis. We use linear interpolation to fill up for the missing values, which constitutes only 0.2% of the total data set. The correlation coefficient between (logarithmic) PV production and power spot prices for 50Hertz, Amprion and TransnetBW are -0.3871 , -0.3323 and -0.3436 , respectively. All the correlation coefficients are negative, which is in line with the intuitive relationship between PV production and power prices. The further analysis will focus only on 50Hertz, as the findings for the three TSO's are rather similar.

Based on the scatter plot in Figure 13 (bottom), which indicates a linear relationship between power spot prices and logarithmic PV production, we propose to model $S(t)$, the electricity spot prices, by

$$S(t) = -\rho \ln PV(t) + R(t). \tag{3.0.1}$$

Here, $R(t)$ is a stochastic process for the power spot prices after the influence of PV production has been accounted for, while the constant ρ is the regression coefficient between spot prices and PV production. The estimated ρ is $\hat{\rho} = 0.3871$. We compute a data series for the process $R(t)$ by appealing to the relationship $R(t) = S(t) + \rho \ln PV(t)$. In Figure 14, we plot the ACF of the residual

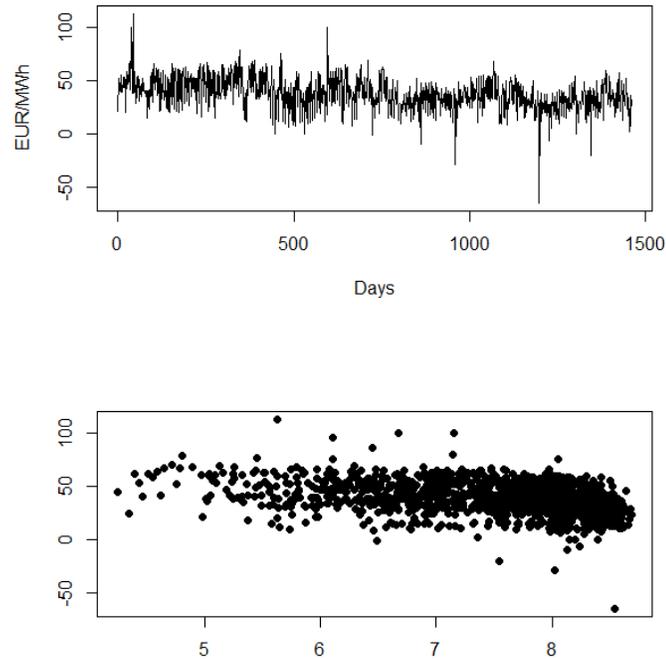


Figure 13: (Left) German Electricity Spot Prices. (Right) Scatter Plot of Spot Prices and $\ln(\text{PV Production})$

data, where the weekend effect is evident.

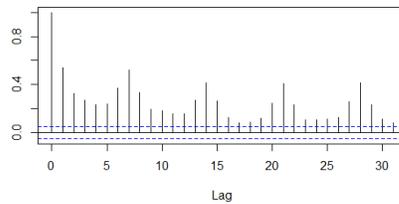


Figure 14: ACF of $R(t)$.

3.1 Modelling the residual process $R(t)$

Assume $R(t)$ to be the composition of a deterministic seasonality function, $U(t)$, and a stochastic term, $Y(t)$, of the form,

$$R(t) = U(t) + Y(t), \quad (3.1.1)$$

where $U(t)$ is the combination between weekends, seasons and yearly cyclical pattern:

$$U(t) = a + bt + \sum_{i=1}^2 l_i W_i + \sum_{j=2}^4 k_j M_j + c \cos\left(\frac{2\pi u(t-v)}{365}\right). \quad (3.1.2)$$

We add dummy variables to the function U to represent weekend and seasonal effects. Here, $W_1 = 1_{t=\text{Saturday}}$ and $W_2 = 1_{t=\text{Sunday}}$, indicating on Saturday and Sunday, respectively. Similarly, M_2, M_3 and M_4 is indicating on the winter, summer and fall seasons, respectively. By including these dummies, we can model the particular effects of different days and seasons on the power price. Many previous studies have included dummy variables to describe the seasonal behavior and predictable component in the electricity prices, for example Lucia & Schwartz [27], Huisman & Mahieu [22], Huisman *et. al* [23] and Higgs & Worthington [21]. Since we include the intercept term in the regression, one indicator variable should be omitted to avoid perfect multi-co-linearity. We eliminated Spring from the regression, which thus becomes a benchmark to the other variables. The parameters of $U(t)$ are estimated by *nonlinear least square* in R and reported in Table 10. We observe that the weekend and season effects are rather pronounced, while the amplitude of the trigonometric term is only $c = 1.918$. The estimated coefficients are all significant with $p < 2e - 16$ except for the last coefficient, v , being the phase shift.

Figure 15 shows our proposed seasonality function in equation (3.1.2) (plotted in red) along with the ACF and PACF of the deseasonalized data. Evidently, the weekend effect is no longer present. The PACF suggests an AR(3)-dynamics for the stochastic process $Y(t)$, and the regression coefficients of this time series model is reported in Table 11. The ACF of the residuals of the AR(3)-process is illustrated in Figure 16, giving reason for zero correlated noise. The K-S statistics $D = 0.032$ with $p\text{-value} = 0.092$ is not significant to reject the normality assumption on the residuals of the AR(3)-process for $Y(t)$, and we therefore stick to the choice of Gaussian independent noise in this case. However, a closer inspection of the density and Q-Q plots (see Figure 16), points toward some heavy tailed behaviour, and one may do a refined modelling of the noise using, for example, the normal inverse Gaussian (NIG) distribution which is rather common in finance (see Barndorff-Nielsen [4]). Worth noticing, is that the normal distribution fits the residual noise rather well, despite the occurrence of spikes in power price data.

Table 10: Estimated parameters of seasonality function, $U(t)$

a	l_1	l_2	k_2	k_3	k_4	b	c	u	v
50.760	-10.340	-18.220	6.380	4.988	7.883	-0.015	1.918	1.069	30.510

Table 11: Regression parameters of AR(3)

β_1	β_2	β_3
0.404	0.086	0.070

4 Application to energy derivatives

In this section, we will discuss some energy derivatives and how our model can contribute in the analysis of these. More specifically, we propose two classes of exotic derivatives, energy quanto options and virtual power plant (VPP) contracts for PV production. Moreover, we include a discussion on continuous-time modelling of PV production and electricity price, which can be convenient in pricing energy derivatives.

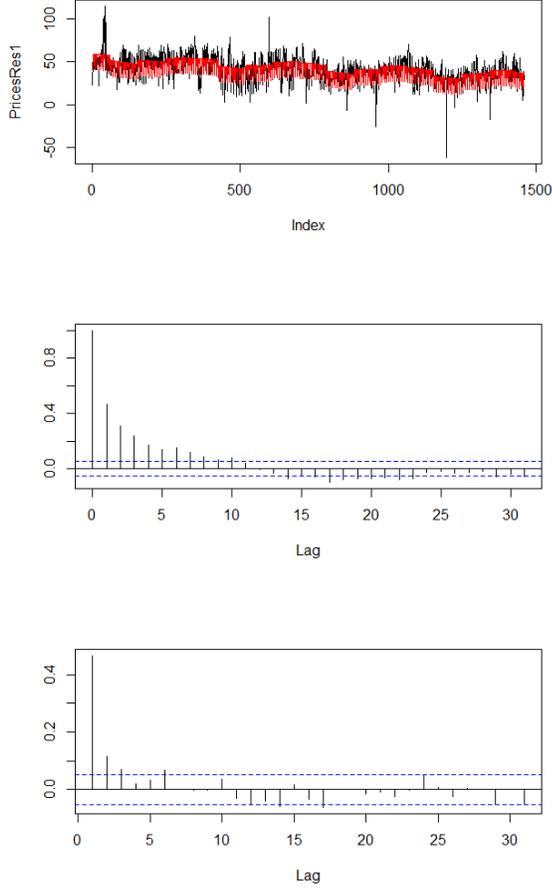


Figure 15: (Top) Fitted plot with seasonality function. (Middle) ACF of deseasonalized data. (Bottom) PACF of deseasonalized data.

4.1 Quanto option

Renewable energy production, such as PV, is intermittent. Over-production cannot be cost-efficiently stored, and any under-production must be compensated for by other power sources, generated by coal, gas or nuclear say. On the other hand, power generation from fossil fuels is highly important for the stability of the supply, but, like in the German market, they can only cover the residual demand left from renewables. As mentioned earlier, sunny days with high PV production will lead to an increase in supply, which will contribute to a decrease in prices (recall the analysis in the previous section). A coal or gas-fired power plant will be hit double in this case, both from less volume and from reduced prices. Hence, to manage their price and volume risk, it would make sense to look for options that can hedge against the joint price and supply risk. For example, a put-call structure on price and PV production could be attractive, having the payoff

$$\max(K - S(T), 0) \times \max(PV(T) - L, 0), \quad (4.1.1)$$

where K and L are contract-specified strikes and $S(T)$ and $PV(T)$ are the power spot price and PV production, respectively. The holder of such a contract will receive a payment $(K - S(T)) \times (PV(T) - L)$ as a compensation for the price and production loss if $PV(T) > L$ (high supply from PV) and

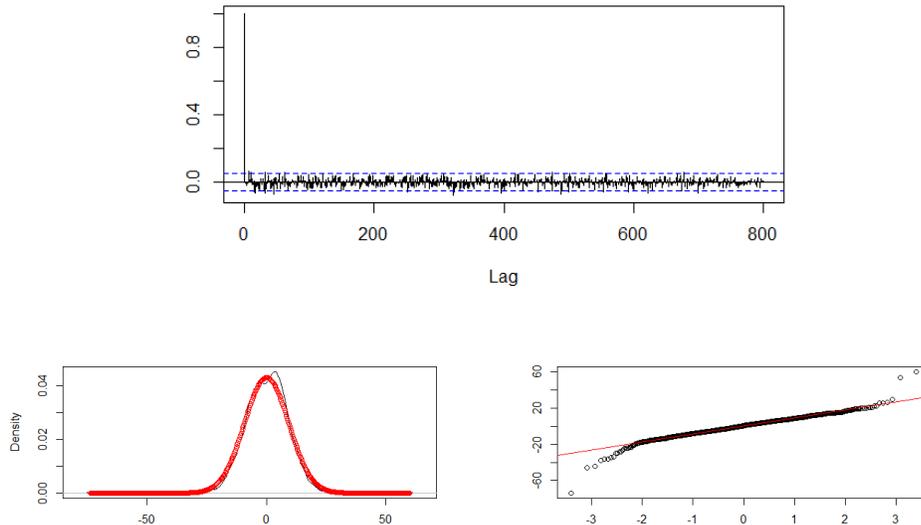


Figure 16: Plots of residuals analysis. (Top) ACF, (Bottom left) Density, (Bottom right) Q-Q plot.

$S(T) < K$ (low prices). This option contract is an example of an *energy quanto option*, extensively discussed in Caporin *et al.* [14], Benth *et al.* [9] and Brik and Roncoroni [12]. To understand the distribution of the payoff from such options, their price and the optimal design of strikes, a joint model for the power price and PV production is required. We have proposed such a model in the previous Sections, suitable for such analysis.

In recent years, quanto options have become popular in weather and energy risk management. The options are typically written on some average of the price and some weather-related variable, mostly a temperature index measuring demand. However, such contracts are attractive also from the supply perspective, then involving a component from wind or sun or precipitation. Quanto options may be much cheaper than a position in two separate plain vanilla options written on energy prices and weather factors, respectively (see Caporin *et al.* [14]). Options on power price exist in many power markets, while derivatives on weather-variables are traded to some extent at the Chicago Mercantile Exchange (see Benth and Šaltytė Benth [8] for an overview). Benth *et al.* [9] present an approach to pricing of such options including a pricing measure, while Brik and Roncoroni [12] study the optimal design of such options.

4.2 PV Virtual Power Plant derivatives

According to Baldi [3], having traditional financial derivatives is not sufficient to mitigate the major price risk in power markets. Therefore, “Tolling agreements” have been created as a complement class of derivatives fulfilling the needs of specific risk sharing in the energy market, particularly targeting the power generation activity. Tolling agreements can be viewed as long-term electricity purchase contracts. In a real world situation, once entering a tolling contract, the toller has to supply the tollee the fuel to generate power, and, in return, the tollee must provide the power to the toller. Basically, the toller will face the price spread risk arising from fuel prices and commercializing the electricity, while the tollee has to endure the technical risks from operations and maintenance of the power plant. An alternative to such physical contracts is virtual power plant (VPP) contracts, being financially equivalent to operating a power plant.

Typically, VPP contracts has been designed for carbon-fueled power generation, and in a simple setting the VPP derivatives become a strip of spread options on the prices of power and gas or coal (spark or black spread, respectively). We refer to Eydeland and Wolyniec [19] for more details. In our context, we propose a VPP contract based on PV production, that is, a financial agreement that gives the holder access to running a PV power plant.

A stylized PV VPP contract can take the following payoff: A PV generator in the German market will paid the spot power price at EEX for her production, but due to subsidies she is guaranteed a minimum positive price K . Holding a PV VPP contract over a time period $[0, T]$ will then yield an income stream with present value

$$\sum_{t=0}^T e^{-rt} \max(S(t), K) PV(t), \quad (4.2.1)$$

where $S(t)$ is the spot price of power and $PV(t)$ is the PV production, both at time t , and $r > 0$ is the interest rate (assumed to be constant here). We can view this contract as a strip of energy quanto options on PV and power price, having payoff $\max(S(t), K) \times PV(t)$. To value such PV VPP options, we must have access to a joint PV production and spot power price model.

4.3 A continuous-time AR(p) dynamics

In mathematical finance, dealing with pricing and hedging of derivatives, continuous-time stochastic processes are frequently used because of their mathematical convenience. In this Subsection, we present a continuous-time process for the autoregressive models occuring in both our proposed PV production times series, and for the power price.

First, let us define the p -dimensional stochastic process $\mathbf{Z}(t)$ for $t \in [0, \infty)$ and $p \in \mathbb{N}$ by,

$$d\mathbf{Z}(t) = A\mathbf{Z}(t) dt + \mathbf{e}_p \sigma(t) dL(t), \quad (4.3.1)$$

where A is the $p \times p$ -matrix

$$A = \begin{bmatrix} 0 & & I \\ -\alpha_p & \dots & -\alpha_1 \end{bmatrix}, \quad (4.3.2)$$

for constants $\alpha_1, \dots, \alpha_p$ being positive and I being the $(p-1) \times (p-1)$ identity matrix. Furthermore, \mathbf{e}_p is the p 'th canonical basis vector in \mathbf{R}^p and L is a real-valued square integrable Lévy process. We suppose that σ is a measurable deterministic function being square integrable on any interval of the positive real line, and thus in particular being integrable with respect to the Lévy process.

Following Brockwell [11], a CAR(p)-process is defined as

$$Z(t) = \mathbf{e}_1^\top \mathbf{Z}(t), \quad (4.3.3)$$

with \mathbf{e}_1 being the first canonical basis vector on \mathbb{R}^p . An explicit representation of Z is

$$Z(t) = \mathbf{e}_1^\top e^{At} \mathbf{Z}(0) + \int_0^t \mathbf{e}_1^\top e^{A(t-s)} \mathbf{e}_p \sigma(s) dL(s), \quad (4.3.4)$$

where $\exp(A)$ is the matrix exponential. From Ch. 4 of Benth and Šaltytė Benth [8], an Euler approximation of CAR(p)-processes on a discrete time scale results in an AR(p)-process. Also, (see Benth and Šaltytė Benth [8]) the regression parameters are linked to the parameters $\alpha_i, i = 1, \dots, p$ in the A -matrix by a linear system of equations. The process $Z(t)$ admits a limiting distribution as $t \rightarrow \infty$ whenever the eigenvalues of A have strictly negative real part. In this case, one typically refers to Z as stationary.

Motivated by the AR(3) time series $X(t)$ in (2.2.1), empirically studied in Subsection 2.3, for the PV production, we can model X in continuous-time by letting $X(t) = Z(t)$, with Z defined in (4.3.3),

for $p = 3$, and the volatility $\sigma(t)$ being defined as in (2.4.1). We recall the standardized residuals to be modelled by a skew normal distribution. Unfortunately, according to Dominguez-Molina and Rocha-Arteaga [16], the skew normal distribution is not infinitely divisible, and is therefore unsuitable as a basis for defining a Lévy process. However, there exists a rich class of (more complex) skewed distributions which are infinitely divisible, for example the skewed Student t (see Dominguez-Molina and Rocha-Arteaga [16]) or the generalized hyperbolic skewed Student t (see Aas and Hobæk Haff [1]). Using one of these, we can define a Lévy process L . A rather crude choice is of course to let $L = B$, a Brownian motion, ignoring the skewness in data. Recalling the autoregressive coefficient β_1, β_2 and β_3 of the AR(3)-dynamics of X in (2.3.2), we can recover α_1, α_2 and α_3 by the system of equations (see Benth and Šaltytė Benth [8])

$$\begin{aligned} 3 - \alpha_1 &= \beta_1 \\ 2\alpha_1 - \alpha_2 - 3 &= \beta_2 \\ \alpha_2 + 1 - (\alpha_1 + \alpha_3) &= \beta_3 \end{aligned}$$

Based on the estimates in Table 5 of X , we find the corresponding α 's by solving the above system. The estimated figures for the deseasonalized PV production process on logarithmic scale are reported in Table 12. We remark that the eigenvalues of the corresponding A -matrices have negative real part, so we conclude with stationarity of the continuous-time model of X .

We do a similar analysis for the process Y in the residual process R of the power price dynamics, presented in equation (3.1.1). Again we have an AR(3)-time series model, but now with noise being Gaussian, arguing in favor of choosing $L = B$, a Brownian motion, and $\sigma(t) = \sigma$, a constant. Based on the estimates of the autoregressive coefficients in Table 11, we derive the corresponding parameters in A reported in Table 13. Also in this case the eigenvalues have negative real part, thus yielding a stationary CAR(3)-dynamics for $Y(t)$.

Table 12: Fitted regression parameter of CAR(3)-PV production

TSOs	α_1	α_2	α_3
50Hertz	2.589	2.076	0.479
Amprion	2.595	2.079	0.480
TransnetBW	2.605	2.103	0.434

Table 13: Fitted regression parameters of CAR(3) - Power Spot Prices

α_1	α_2	α_3
2.596	2.107	0.440

The pricing of quanto options will involve computing an expectation $\mathbb{E}_Q[f(S(T))g(PV(T))]$, for some given functions f and g (call and/or put payoff functions, say). The expectation is taken with respect to some pricing measure Q being equivalent to P . From the models that we have proposed for S and PV , and with applying the continuous-time process CAR(3) version with explicit representation in (4.3.4), we face the problem of computing

$$\mathbb{E}_Q \left[h \left(\int_0^T \mathbf{e}_1^\top e^{A_{PV}(T-s)} \mathbf{e}_3 \sigma(s) dL(s), \int_0^T \mathbf{e}_1^\top e^{A_S(T-s)} \mathbf{e}_3 \sigma dB(s) \right) \right], \quad (4.3.5)$$

for some function h depending on the payoff structure defined by f and g . Here, A_{PV} is the A -matrix for the PV production process (e.g., the X -process), and A_S for the power spot price (e.g. the Y -process). To obtain a pricing measure, we need to have some idea about the risk premium in the market, which is a delicate issue that we will not touch here. However, one may produce

a rather flexible class of risk premia by using a specification of the market price of risk, obtained via Girsanov transformation in the Brownian case and Esscher transform for the Lévy process (see Benth *et al.* [6]). The computation of the expectation functional can be performed using a Fourier transform approach. We refer to Benth [5] for more details on this for quanto options.

Admittedly, we need a continuous-time version of the seasonality functions for both PV production and power price. The latter involves some indicator functions for specific days, which can be trivially extended. A more refined analysis, however, will take into account the intraday price pattern. For the PV intensity function, this will be a continuous function of time, and we can extend it for other times as well, including night time when it is set to zero. These issues will be treated in more detail in a future study.

5 Conclusion

We have proposed a novel model for the (logarithmic) PV production dynamics using the sun intensity to explain the seasonal mean variation. Our empirical analysis is based on four years of daily 12pm production data from three transmission system operators in Germany. The deseasonalized (logarithmic) PV production data followed nicely an autoregressive process of order 3, with seasonal volatility and non-Gaussian noise. We suggested to model the noise by the skewed normal distribution.

We studied some applications of our PV model. First, we built a relationship between electricity spot prices and logarithmic PV production by using PV production as an explanatory variable in the power price dynamics. We had available daily 12pm German power prices, and showed that there is a negative relationship as expected. After removing the influence of PV production, the power spot price dynamics could be modelled very well using a simple autoregressive dynamics of order 3, with normally distributed noise.

As a further application, we discussed quanto options and PV virtual power plant contracts, and how our model can be applied to analyse these. In particular, we showed that one can develop continuous-time models for our dynamics, so-called continuous-time autoregressive processes. Such processes are convenient when pricing quanto options, but also when assessing the joint price and production risk and when hedging such options. Our model and analysis can be useful for both renewable power producers and those generating power using fossil fuels. Both sides face significant supply and price risk, where various exotic option structures may be attractive for hedging purposes.

In future studies we want to analyse the price dynamics of PV production for all hours during day time. This could be combined with a model for power prices, where one may expect different impact from PV over the daytime. It is also of interest to create models including wind power production as well. We will return to these challenging problems in a future study.

Acknowledgements

The authors thank Prof. Dr. Rüdiger Kiesel and Nikolaus Luckner from University of Duisburg-Essen for providing electricity spot price data. Financial support from FINEWSTOCH, a research project funded by the Norwegian Research Council, is acknowledged.

References

- [1] Aas, K., Hobæk Haff, I. (2005). The generalised hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, **4**(2).
- [2] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, pp. 171–178.
- [3] Baldi, F. (2010). Switch, switch, switch! A regime-switching option-based model for valuing a tolling agreement. *The Engineering Economist*, **55**(3), pp. 268–304, DOI: 10.1080/0013791X.2010.504807.
- [4] Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance & Stochastics*, **2**(1), pp. 41–68.
- [5] Benth, F. E. (2016). Cointegrated commodity markets and pricing of derivatives in a non-Gaussian framework. In *Advanced Modelling in Mathematical Finance. Festschrift in honour of Ernst Eberlein*. J. Kallsen and A. Papapantoleon (eds.). Springer Verlag, pp. 477–496.
- [6] Benth, F. E., Šaltytė Benth, J., Koekebakker, S. (2008). *Stochastic Modelling of Electricity and Related Markets*, World Scientific, Singapore.
- [7] Benth, F. E., Šaltytė Benth, J. (2011). Weather derivatives and stochastic modelling of temperature. *International Journal of Stochastic Analysis*, Vol 2011, Article ID 576791, 21 pages.
- [8] Benth, F. E., Šaltytė Benth, J. (2012). *Modelling and Pricing in Financial Markets for Weather Derivatives*, World Scientific, Singapore.
- [9] Benth, F. E., Lange, N., and Myklebust, T. A. (2015). Pricing and hedging quanto options in energy markets. *Journal of Energy Markets*, **8**(1), pp. 1–35.
- [10] Benth, F. E., Taib, C. M. I. C. (2013). On the speed towards the mean for continuous time autoregressive moving average processes with applications to energy markets. *Energy Economics*, **40**, pp. 259–268.
- [11] Brockwell, P. J. (2001). Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics*, **53**(1), pp. 113–124.
- [12] Brik, R. I., and Roncoroni, A. (2016). Static mitigation of volumetric risk. *Journal of Energy Markets*, **9**(2), pp. 111–150.
- [13] Bolancé, C., Guillen, M., Pelican, E., Vernic, R. (2008). Skewed bivariate models and nonparametric estimation for the CTE risk measure. *Insurance: Mathematics and Economics* **43** (3), pp. 386–393.
- [14] Caporin, M., Preš, J., Torro, H. (2012). Model based Monte Carlo pricing of energy and temperature Quanto options. *Energy Economics*, **34**, pp. 1700–1712.
- [15] Cococcioni, M., D'Andrea, E., Lazzarini, B. (2011). 24-Hour-ahead forecasting of energy production in solar PV systems. *International Conference on Intelligent Systems Design and Applications, ISDA*, **6**, pp. 1276–1281.

- [16] Dominguez-Molina, J. A., Rocha-Arteaga, A. (2007). On the infinite divisibility of some skewed symmetric distributions. *Statistics & Probability Letters*, **77**, pp. 644–648.
- [17] Duffie, J.A., Beckman, W.A. (2013). *Solar Engineering of Thermal Processes*, 4th ed, Wiley & Sons..
- [18] Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, **51**, pp. 239–248.
- [19] Eydeland, A., and Wolyniec, K. (2003). *Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging*. Wiley & Sons.
- [20] Härdle, W., Lopez Cabrera, B. (2012). The implied market price of weather risk. *Applied Mathematical Finance*, **19**(1), pp. 59–95.
- [21] Higgs, H., Worthington, A. C. (2005). Systematic features of high frequency volatility in the Australian electricity market: Intraday patterns, information arrival and calendar effects. *The Energy Journal*, **26**(4), pp. 23–41.
- [22] Huisman, R., Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics*, **25**, pp. 425–434.
- [23] Huisman, R., Huurman, C., Mahieu, R. (2007). Hourly electricity prices in day-ahead markets. *Energy Economics*, **29**, pp. 240–248.
- [24] Hocaoglu, F. O. (2011). Stochastic approach for daily solar radiation modeling. *Solar Energy*, **85**, pp. 278–287.
- [25] Honsberg, C., Bowden, S. (2016). Photovoltaic education network. [Available at] www.pveducation.org.
- [26] Khalil, H. B., Zaidi, S. J. H. (2014). Energy crisis and potential of solar energy in Pakistan. *Renewable and Sustainable Energy Reviews*, **31**, pp.194–201.
- [27] Lucia, J. J., Schwartz, E. S. (2002). Electricity prices and power derivatives: Evidence from the Nordic Power Exchange. *Review of Derivatives Research*, **5**, pp. 5–50.
- [28] Veraart, A. E. D and Zdanowicz, H. (2016). Modelling and predicting photovoltaic power generation in the EEX Market. Available at SSRN: <https://ssrn.com/abstract=2691906> or <http://dx.doi.org/10.2139/ssrn.2691906>.
- [29] Yao, J. (2014). Determining the energy performance of manually controlled solar shades: A stochastic model based co-simulation analysis. *Applied Energy*, **127**, pp. 64–80.