Humboldt-Universität zu Berlin

School of Business and Economics

Forecasting of Solar Power Generation

Author:
David Schulte
615663

Supervisors:
Prof. Dr. López Cabrera

Seminar Paper for the course
Statistical Tools in Finance and Insurance

March 10, 2022

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Abstract

Statistical Tools in Finance and Insurance

School of Business and Economics

Forecasting of Solar Power Generation

by David Schulte

In this paper, I am introducing a method for forecasting solar power generation in Germany. The method consists of transforming time series data, detrending it, removing the seasonality component, and finally using an ARMA model to predict future values. Chapter 1 will be an introduction. In Chapter 2, we will introduce previous work concerning the topic. In Chapter 3, I will present the data used in this approach. Chapter 4 is the core of this paper, as it includes the analysis and the conception of the proposed model. At last, Chapter 5 will outline the future outlook.

For our analysis, we will use Python 3.9 and the statsmodels library.

Contents

Abstract **i**

1 Introduction **1**

2 Literature Review **2**

3 Data **3**

4 Analysis **4**
4.1 Transformation 4
4.2 Seasonality 5
4.3 ARMA 7

5 Conclusion **11**

6 Bibliography **12**

Chapter 1

Introduction

Countries all across the globe are increasing the amount of renewable energies in their power generation sources. One of the main components of this new strategy is solar power. The technology of solar panels that are essential for this shift towards renewable energy has become more and more sophisticated. At the same time, more panels are being installed. For example, in Germany, the power generated by solar power per year has been increased by a factor of 856 between the years 2000 and 2020. Like all kinds of renewable energies, the generation of solar power cannot be controlled, as it depends on solar radiation, temperature, and other factors. Since the amount of energy generated greatly influences the energy price, variation in solar power generation results in financial risk. The goal of this paper is to quantify this risk. This is especially important for investors and industries that heavily depend on stable energy prices, in order to find appropriate ways to hedge against the imposed risk. I will model the daily amount of solar power generation in Germany, and analyze the results.

Chapter 2

Literature Review

As the usage of renewable energies rapidly increased since the beginning of the 21st century, the modeling of weather and energy generation using renewable sources has become more important in the field of finance.

Alaton et al. (2002) modeled temperature data and used the market price of risk to value weather derivatives on the temperature in Stockholm. Campbell and Diebold (2005) focused on the distribution of temperature and approximated densities for the temperature in multiple cities in the United States.

Reikard (2009) used an ARIMA model to make short-term forecasts of the intensity of solar radiation.

Mitrentsis and Lens (2022) used Machine Learning models trained on a set of features including temperature, humidity, wind speed and radiation, to make accurate and interpretable short term predictions of solar power generation.

Härdle et al. (2021) used a Gaussian CARMA model of wind power utilization to prize wind power futures in Germany. This paper is heavily influenced by their work, as the modeling process is a simplified version of their approach.

What distinguishes this work, is that we will not model weather data but the solar power generation. We will not rely on other data sources than historical generation data provided by German transmission system operators. Therefore, our model does not only incorporate meteorological processes but also the increase of utilized solar panels in Germany.

Chapter 3

Data

For this project, we will use the daily solar energy generation in Germany in the years from 2010 until 2020. The data was collected by Fraunhofer Institute for Solar Energy Systems. The daily values are the averages of the corresponding values per day reported by Germany's four transmission system operators 50Hertz, Amprion, TenneT and TransnetBW.
The data can be accessed at: https://energy-charts.info/

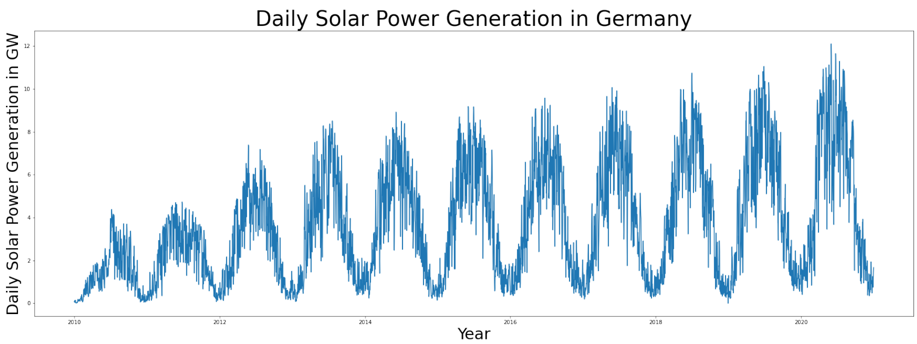

FIGURE 3.1: Dataset

Chapter 4

Analysis

4.1 Transformation

The first thing, that stands out, when taking a look at **??**, is that there is a very strong seasonal component in the data. However, this is not the first step in our preparation process. First, we will examine the distribution of our daily values and try to transform our data such that the distribution has more similarity to a normal distribution.

$$\tilde{U}_t = \frac{U_t - U_{\min}}{U_{\max} - U_{\min}} \tag{4.1}$$

$$U_t^* = \log(\tilde{U}_t / (1 - \tilde{U}_t)) \tag{4.2}$$

To assure that our second transformation is well defined, we add/subtract a small ϵ value to the minimum/maximum value of $\tilde{U}$, such that all values are in the interval $(0, 1)$. Note that both transformations are bijective, which allows us to do an inverse transformation after modeling.

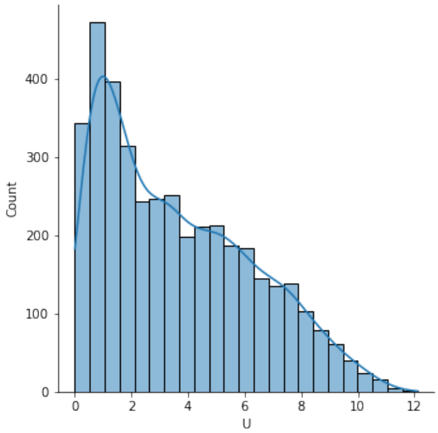

FIGURE 4.1: Distribution of the observed data

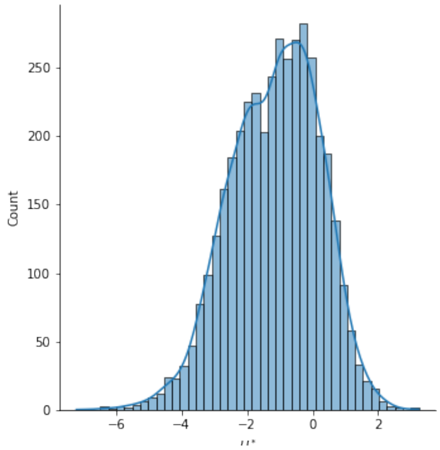

FIGURE 4.2: Distribution of the transformed data

As we can see, the result is not perfect but noticeably closer to a normal distribution, which justifies the next steps in our transformation.

4.2 Seasonality

We will now try to decompose the transformed data into intercept, trend, seasonality and further variation. To do so, we will apply a linear regression model with artificially constructed regressors.

$$U_t^* = \overbrace{\beta_0}^{\text{Intercept}} + \overbrace{\beta_1 \cdot t}^{\text{Trend}} + \overbrace{\beta_2 \cdot \sqrt[4]{t} \cos\left(2\pi \frac{(t-11)}{365}\right)}^{\text{Seasonality}} + \overbrace{X_t}^{\text{Residuals}} \tag{4.3}$$

While the interpretation of intercept and trend is obvious, the artificial regressor for seasonality requires further explanation. We model the yearly seasonality by a cosine function with a period of exactly one year. We expect the low of our solar power generation to be on average around Winter solstice, the shortest day of the year. This day is exactly 11 days before New Year, which is the beginning of our time series data per year. That is why we shift the cosine wave by that amount. The second component that comes into play is that the aptitude of our cosine function increases over the years. This increase is not linear. After trying different terms, I ended up using the 4^{th} square root as an increasing factor.

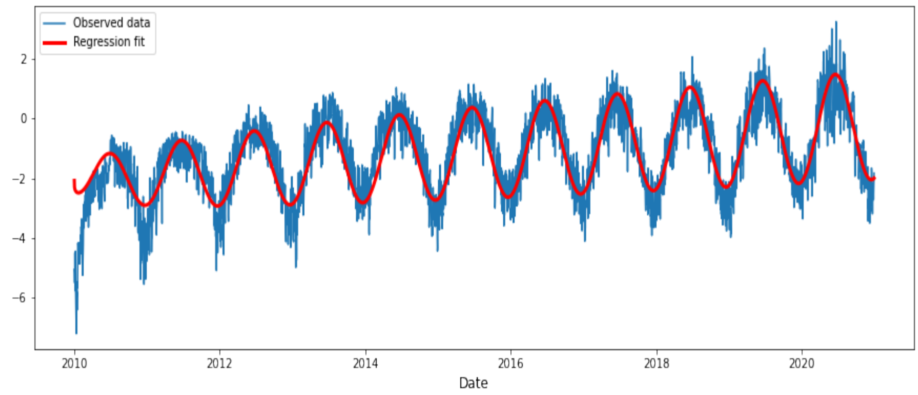

FIGURE 4.3: Seasonality of the data

Our linear regression computes the following parameters. One has to be careful, when interpreting these results, as we apply the regression on our transformed time series, and not the initial one.

Dep. Variable:	y	**R-squared:**	0.724
Model:	OLS	**Adj. R-squared:**	0.724
Method:	Least Squares	**F-statistic:**	5272.
Date:	Thu, 10 Mar 2022	**Prob (F-statistic):**	0.00
Time:	21:09:03	**Log-Likelihood:**	-4394.5
No. Observations:	4018	**AIC:**	8795.
Df Residuals:	4015	**BIC:**	8814.
Df Model:	2		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
beta0	-2.0841	0.023	-91.413	0.000	-2.129	-2.039
beta1	0.0005	9.83e-06	46.923	0.000	0.000	0.000
beta2	-0.2282	0.002	-92.181	0.000	-0.233	-0.223

Omnibus:	507.081	**Durbin-Watson:**	0.620
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1086.897
Skew:	-0.766	**Prob(JB):**	9.62e-237
Kurtosis:	5.036	**Cond. No.**	4.64e+03

Our trend coefficient β_1 quantifies the trend in our data. It most likely originates in the continuing development of solar panels in Germany.

All our coefficients are statistically significant. Moreover, the high value for R^2 is remarkable. It proves that most variation of the data originates from its trend and seasonality.

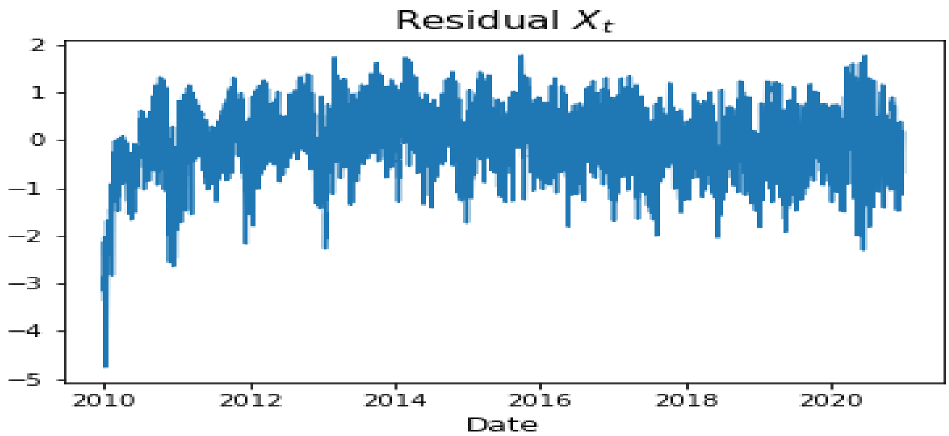

FIGURE 4.4: Residuals of the regression

We have outliers in our residuals at the beginning of our time series. The reasons are both the transformation, in which the minimum value becomes very small after being transformed, as well as the square-root term in our model. Since the residuals seem reasonable later in the series, we will resort to a simple trick, namely dropping the first year (2010) from our data set from now on. We can easily do this, as we are interested in future predictions. Also, we have an abundance of data in relation to the complexity of our model.

4.3 ARMA

By now, we have determined the strong long-term signals in our data. From now on, we will focus on short-term correlations. We will inspect the partial autocorrelation function.

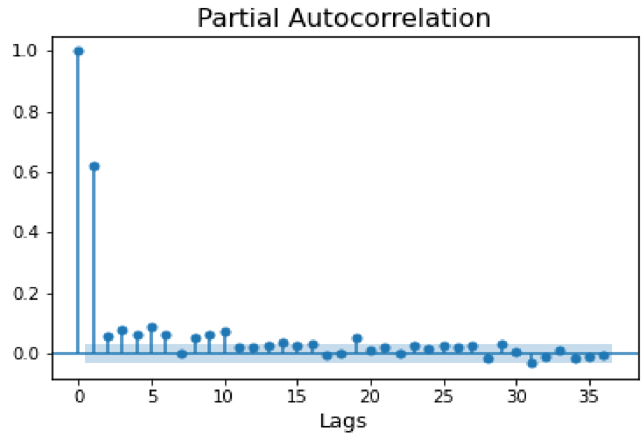

FIGURE 4.5: PACF of the residuals

We have very strong autocorrelation for one lag and significant values for up to 10 lags. Based on this insight, a model using either 1 or 10 lags seems suitable. After trying out both options, I decided upon the one using only one lag. The residuals for both models are of similar size, and following the principle of Occam's razor, we prefer the simpler model. Our ARMA(1,1) model has the following form.

$$X_t = c + \phi_1 X_{t-1} + \theta_i \epsilon_{t-1} + \epsilon_t \tag{4.4}$$

We inspect the fit of the model. These values are computed by making a one-step-ahead in-sample prediction.

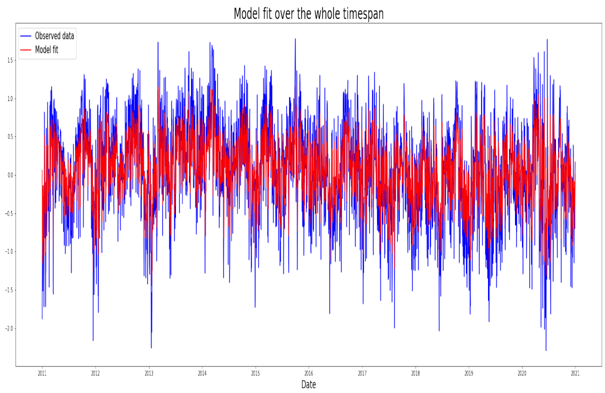

FIGURE 4.6: Model fit

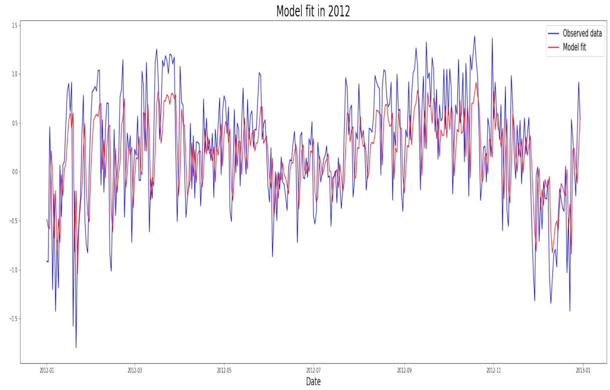

FIGURE 4.7: Model fit in detail

The statistics of our model are summarized in the following.

Dep. Variable:	cleaned	**No. Observations:**	3653
Model:	ARIMA(1, 0, 1)	**Log Likelihood**	-2738.148
Date:	Thu, 10 Mar 2022	**AIC**	5484.296
Time:	21:09:06	**BIC**	5509.109
Sample:	01-01-2011 - 12-31-2020	**HQIC**	5493.132

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0593	0.025	2.338	0.019	0.010	0.109
ar.L1	0.6961	0.018	38.043	0.000	0.660	0.732
ma.L1	-0.1208	0.024	-5.031	0.000	-0.168	-0.074
sigma2	0.2621	0.006	43.612	0.000	0.250	0.274

Ljung-Box (L1) (Q):	0.14	**Jarque-Bera (JB):**	95.96
Prob(Q):	0.71	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.20	**Skew:**	-0.37
Prob(H) (two-sided):	0.00	**Kurtosis:**	3.30

We can observe that all coefficients are statistically significant. Especially interesting is the value of of σ^2(sigma2), which is the estimated variance of our residuals after apply the ARMA model.

In the following, we will inspect the distribution of our residuals.

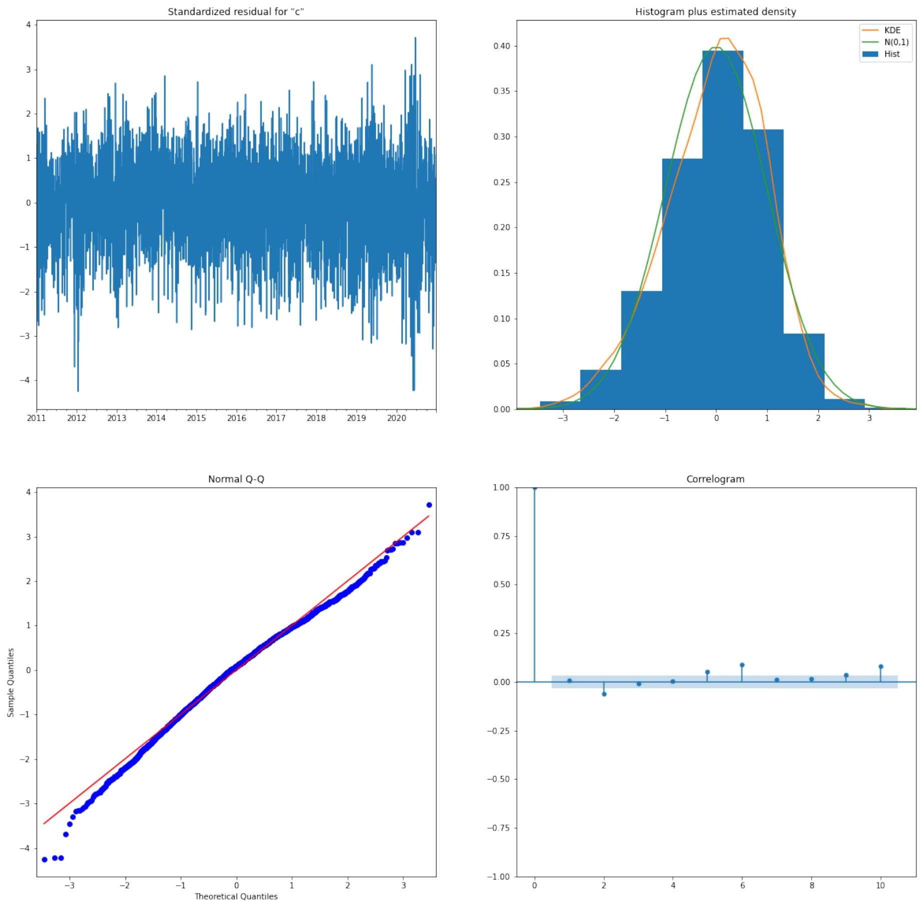

FIGURE 4.8: Properties of the ARMA residuals

The residuals are nearly normally distributed, with deviations in the left tail of the distribution. Furthermore, we can observe that the residuals show very low autocorrelation, further supporting our assumption that an ARMA(1,1) is sufficient.

Chapter 5

Conclusion

We have created a model that explains a large part of the variation in our data, where we exploited the very strong seasonality of solar power generation. After fitting our model, we can also quantify the variation, it can not explain. This variation is especially interesting, as it can be interpreted as a risk factor. Thus, it can be used, when pricing derivatives in the future.

One has to acknowledge that one of the strong points of our model, namely that it depends solely on past solar power generation data, is also its biggest weakness. Especially, when making short-term predictions, the incorporation of real-time meteorological data, as well as, weather forecasts is very promising. Furthermore, one could differentiate between the different regions of the transmission system operators, instead of aggregating them to the power generation of the whole country.

Overall, the results, achieved by our model are promising and have the potential to be used as one component in the pricing of derivatives or strategies involving power trading.

Chapter 6

Bibliography

Alaton, Peter, Boualem Djehiche, and David Stillberger (2002) "On modelling and pricing weather derivatives," *Applied mathematical finance*, 9 (1), 1–20.

Campbell, Sean D and Francis X Diebold (2005) "Weather forecasting for weather derivatives," *Journal of the American Statistical Association*, 100 (469), 6–16.

Härdle, Wolfgang Karl, Brenda Lopez Cabrera, and Awdesch Melzer (2021) "Pricing wind power futures," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70 (4), 1083–1102.

Mitrentsis, Georgios and Hendrik Lens (2022) "An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting," *Applied Energy*, 309, 118473.

Reikard, Gordon (2009) "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar energy*, 83 (3), 342–349.