Karlsruhe Institute of Technology

Department of Economics and Management

Institute of Operations Research (IOR)

Analytics and Statistics

Prof. Dr. Oliver Grothe

Bachelor thesis

Stylized Facts of Macroeconomic and Wind Power Forecast Errors

Author

Tobias Kuhlmann

Matr.Nr.: 1767742

Wirtschaftsingenieurwesen, B.Sc.

Date of submission

12.05.2017

Contents

Abbreviations

ARCH autoregressive conditional heteroscedasticity

ARIMA autoregressive integrated moving average

cdf cumulative distribution function

ECB European Central Bank

GARCH generalized autoregressive conditional heteroscedasticity

iid independent and identically distributed

OLS ordinary least squares

pdf probability density function

RMSE Root mean squared error

sMAPE Symmetric mean absolute percentage error

SPF Survey of Professional Forecasters

TSOs transmission system operators

TSO transmission system operator

Philadelphia Fed Federal Reserve Bank Philadelphia

VAR vector autoregressive

VECM vector error correction model

1 Introduction

The development and estimation of accurate forecasts are of special interest in a lot of industries and in daily life. For politics and economy it is important to know macroeconomic developments in advance, to determine their strategy. As an investor it is essential to predict market movements or interest rates, and for a good energy supply management it is necessary to know rewenable energy's output in advance. *Xcel Energy Inc.*, for example, a power company serving millions of electricity customers in the United States, reduced its forecast error rate by 35%, which enabled a more efficient management of coal and natural gas plants. This has saved more than 60mio.$ to this day. (Xcel Energy Inc. 2017a; Xcel Energy Inc. 2017b).

To improve the accuracy of forecasts it is useful to obtain knowledge on the forecast error. While previous research mainly concentrates on performance evaluation and on statistical properties of actual time series, this work is about statistical properties of the forecast error. The goal is to find *stylized facts of forecast errors*. *Stylized facts* are simplified abstractions of common statistical properties and were first introduced by Kaldor (1961). Cont (2001) defines stylized facts in the scope of asset returns as statistical properties which are common across "many instruments, markets and time periods". Transferred to this work we look for statistical properties which are common across many forecast methods, sectors and time periods. Furthermore, we distinguish between sectors, as there can be properties which are only common in a single sector. In order to do this we will gather characteristics and properties of our time series and look for common denominators.

Furthermore, this work presents the principle of cointegration and error correction models and its classification in the scope of *stylized facts of forecast errors*.

The thesis is structured as follows. After an introduction of stylized facts in section two, section three defines theoretical principles. Section four gives a detailed literature review about stylized facts and statistical analysis of forecast errors, as well as cointegration and error correction. In section five, Data sets and data preprocessing are presented, and then we analyze and gather characteristics of the selected time series. In section six, the conclusion, results are summarized and critically assessed.

2 The Concept of Stylized Facts

The concept of 'stylized facts' was first introduced by Kaldor (1961) when he described that the basic requirement of any economic model should be its capability of explaining characteristic features of an economic process. This can be done with such called stylized facts, which are a simplifying abstraction of statistical properties. They concentrate on broad tendencies and ignore individual detail.

In the scope of asset returns Cont (2001) defines stylized facts as a set of properties which are common across many instruments, markets, time periods and confirmed by independent studies. Cont recognizes that seemingly random asset prices share some statistical properties. Those are obtained by looking for a common denominator among properties of various studies. In rather focusing on generality than precision, Cont summarizes a set of stylized facts. They are usually formulated in terms of qualitative properties, thus Cont is using non-parametric models to analyze data (Cont 2001).

3 Theoretical Principles

Necessary theoretical concepts and notations are described in the following section. First, we explain the basic terms of forecasting, before stationarity, cointegration and error correction are defined. The concepts in this section are fundamental for the analysis in section 5.2.

3.1 Basic Concepts of Forecasting

Forecasting is about making predictions of future values with information available today. The process includes an analysis of past and present values, trends, seasonal effects and many more.

Let y_t denote the last available value and $y_{t+k|t}$ the forecasted value, then the period of time k is said to be the forecast horizon. In other words, it is the period of time between the estimated and actual period of a value. For example, a forecast value estimated 24 hours in advance has a forecast horizon of 24 hours. The forecasting process gets more complicated with higher forecast horizons, as there is less information available at period t.

In order to draw conclusions from forecast errors it is important to understand how forecasts are created. We present a intuitive standard method and a short explanation how forecasting in practice works.

The *persistence model*, also called no-change model, is a simple approach and the baseline model against which other methods are typically compared. It is also part of several forecast error measures. The definition is shown in equation 1, where $y_{t+k|t}$ is the forecasted value, y_t is the last actual value and k the forecast horizon.

$$y_{t+k|t} = y_t \tag{1}$$

Simply put, the forecasted value is just the last actual value (B.-M. Hodge and Michael Milligan 2011).

The forecast models used in practice are often a combination of several and more complex methods. For example, wind power forecasts models are usally build with wind speed forecasts. German transmission system operator (TSO) 50Hertz uses a combination of forecasts from various research institutes. (50Hertz Transmission GmbH 2017). A few of the forecast methods used in the M3 challenge and in the European Central Bank (ECB) Survey of Professional Forecasters (SPF) are autoregressive integrated moving average (ARIMA) models, artificial neural networks and exponential smoothing techniques (Makridakis and Michele Hibon 2000; Meyler and Rubene 2009).

3.2 Forecast Error

A forecast error is typically defined as the difference between actual and predicted value. For different forecast horizons k the forecast error e is defined as

$$e_{t+k|t} = y_{t+k} - y_{t+k|t} \tag{2}$$

where y_{t+k} is the actual value at time $t+k$ and $y_{t+k|t}$ is the forecast value at time $t+k$ made with information available at period t (Madsen et al. 2005; Bludszuweit, Domínguez-Navarro, and Llombart 2008).

3.3 Stationarity

Stationarity is an important requirement in statistical analysis to be able to identify underlying statistical properties of the data (Cont 2001). It is essential for forecasting methods to estimate future values based on past values' properties and stochastic processes. A time series is *stationary* if its statistical properties are constant over time. A further distinction exists between strictly and weakly stationary. A *strictly stationary* series has all statistical properties constant over time. That is, when all time shifted sequences of a series have the same joint distributions. If it only applies to mean, variance and covariance it is said to be *weakly stationary* (Ruppert 2011a). Whenever the term *stationary* appears in this work, it is assumed to be *weakly stationary.*

If a time series is non-stationary, but the series of differences of one to the next value is stationary, it is integrated of order 1, or $I(1)$. It is integrated of order d, or $I(d)$, if both time series become stationary through differencing d times. (Engle and Granger 1987)

3.4 Cointegration and Error Correction

The concept of cointegration arised from Granger (1981) and Granger and Newbold (1974) in the context of spurious correlations as another way of making time series stationary. Broadly speaking, spurious correlation is present when two independent, non-stationary time series show significant correlation (Ruppert 2011b; Granger and Newbold 1974). If two or more time series have a similar behaviour and so a similar trend, we can use this relation for detrending, thus making a time series stationary.

The presented models and notations in this section are from Kilian and Lütkepohl (2016), Ruppert (2011c), and Johansen (1991). Let $y_{1,t}$ and $y_{2,t}$ be two non-stationary, $I(d)$ time series. y_1 and y_2 are now cointegrated if both are $I(1)$ and there is a linear combination $y_{1,t} - \lambda * y_{2,t}$ of the initial non-stationary series which is stationary. A common trend model is presented in equations 3 and 4.

$$y_{1,t} = \beta_1 W_t + \epsilon_{1,t} \tag{3}$$

$$y_{2,t} = \beta_2 W_t + \epsilon_{2,t} \tag{4}$$

where β_1 and β_1 are nonzero coefficients, W_t is the trend of both series and $I(1)$ and $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are noise processes and $I(0)$. If a linear combination like in equation 5 with $\lambda = \beta_1/\beta_2$ exists, the linear combination series is detrended and thus stationary.

$$\beta_2(y_{1,t} - \lambda y_{2,t}) = \beta_2 y_{1,t} - \beta_1 y_{2,t} = \beta_2 \epsilon_{1,t} - \beta_1 \epsilon_{2,t} \tag{5}$$

This definition can also be extended to multivariate data with more than two time series.

Cointegration only models a long-term relationship and not short-term dynamics. To take this into account, *error correction models* have been developed. An error correction model uses a cointegration relationship to model and correct the linear combinations short-term deviation from its mean, called the "error". Simply put, it corrects the last period's error in the current period. The idea of an error correction model for two time series is shown in equations 6, 7.

$$\Delta y_{1,t} = \phi_1(y_{1,t-1} - \lambda y_{2,t-1}) + \epsilon_{1,t} \tag{6}$$

$$\Delta y_{2,t} = \phi_2(y_{1,t-1} - \lambda y_{2,t-1}) + \epsilon_{2,t} \tag{7}$$

In equation 6 and 7, $\Delta y_{1,t}$ and $\Delta y_{2,t}$ are the first difference time series, the coefficients ϕ_1 and ϕ_2 specify the speed of mean-reversion and $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are noise processes. $(y_{1,t-1} - \lambda y_{2,t-1})$ is the last periods's deviation of the linear combination from its long-term mean. Equation 8 is the vector notation of equations 6 and 7, with the vectors defined in equation 9. β is called the cointegration vector and α the adjustment matrix.

$$\Delta Y_t = \alpha\beta^T Y_{t-1} + \epsilon_t \tag{8}$$

$$\Delta Y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}, \epsilon_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \alpha = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \beta = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \tag{9}$$

Equation 10 is the result of subtracting equation 7 λ times from equation 6.

$$\Delta(y_{1,t} - \lambda y_{2,t}) = (\phi_1 - \lambda\phi_2)(y_{1,t-1} - \lambda y_{2,t-1}) + (\epsilon_{1,t} - \lambda\epsilon_{2,t}) \tag{10}$$

The term $(\phi_1 - \lambda\phi_2)(y_{1,t-1} - \lambda y_{2,t-1})$ from equation 10 causes the error correction. A negative coefficient $(\phi_1 - \lambda\phi_2)$ brings a positive correction $\Delta(y_{1,t} - \lambda y_{2,t})$ in a current period t, if the deviation $(y_{1,t-1} - \lambda y_{2,t-1})$ of the last period $t-1$ is negative. If the deviation is positive, the coefficient brings a negative correction. This model can be generalized in vector form for more than two time series, which is shown in equation 11, where $\Pi = \alpha\beta^T$, $\Gamma_1\Delta Y_{t-1} + ... + \Gamma_{p-1}\Delta Y_{t-p+1}$ are weighted lagged values of the series, μ is a mean vector and D_t a vector of seasonal dummies.

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1\Delta Y_{t-1} + ... + \Gamma_{p-1}\Delta Y_{t-p+1} + \mu + \Phi D_t + \epsilon_t \tag{11}$$

A detailed description of the multivariate VECM and how to rewrite and extend equations 8 and 9 to equation 11 goes beyond the scope of this work, which is why we refer to Kilian and Lütkepohl (2016) and Johansen (1991) for further reading.

There are two approaches to estimate the coefficients: Engle and Granger (1987) use ordinary least squares (OLS) regression and Johansen (1991) uses maximum likelihood estimation. The *Engle-Granger two-step approach* has its limitations: Pretesting stationarity of the variables, one variable being dependent and only one cointegration vector while there could be several. These limitations are taken into account in *Johansen's approach*. An estimation with the Johansen approach is shown later in section 5.2.6.

4 Literature Overview

This section gives a literature overview on forecast errors in wind power and economics as well as cointegration and error correction models.

4.1 Wind Power Forecast Errors

In recent years there has been a lot of research in the field of wind power forecasts since it is important to anticipate variations in its power production to manage energy supply and prevent outages (H Holttinen, Nielsen, and G Giebel 2002). Most of the papers focus on the probability density function (pdf) of the forecast errors and use wind power data from local transmission system operators (TSOs). Typically only real values are utilized and the forecast is self-produced, in most cases with the persistence method. Bludszuweit, Domínguez-Navarro, and Llombart (2008) analyze forecast errors of such simulated persistence forecasts. The pdf is found to be fat-tailed with a kurtosis from 3 to over 10, which means a pdf with more pronounced peaks, slighter shoulders and heavier tails. The best fitted distribution was the Beta pdf, although it does not fit forecasts up to 24h, as it is not fat-tailed enough. High costs would be involved if large values in the tails are underestimated, as they have the highest economic impact. S Bofinger, A Luig, and H. Beyer (2002), Tewari, Geyer, and Mohan (2011), Armin Luig, Stefan Bofinger, and H. G. Beyer (2001), and Pinson and Kariniotakis (2004) use a similar approach with the beta pdf. Lange (2005) investigates the uncertainty of wind power forecasts and its relationship to wind speed forecasts. The Root mean squared error (RMSE) is decomposed in amplitude and phase errors, which both increase over forecast horizon. Amplitude errors can be balanced while phase errors cannot easily be calibrated out (Lange 2005). B.-M. Hodge and Michael Milligan (2011) use a persistence forecast to model the error pdf over multiple timescales. It is found to be non-Gaussian with a varying kurtosis over timescale. The Cauchy pdf is proposed and compared to a set of other leptokurtic distributions. B.-M. Hodge, Ela, and Michael Milligan (2011) use data from the Electric Reliability Counsil of Texas and Xcel Energy Colorado territory for hourly and day-ahead forecasts. RMSE is increasing over timescale. The shape of the pdf of operational systems is compared with a persistence forecast and found to be significantly different. The shapes depend on the wind park setup and on the timescale. Kurtosis is decreasing and a weak trend to more symmetric distributions over timescale. That means the skewness converges weakly to zero. The Cauchy distribution is proposed to model the error pdf. Suggested topics for further research are analyzing more sophisticated forecasts like ARIMA and incorporate wind power data from other geographic areas. B.-M. Hodge, Lew, et al. (2012) come to the same conclusions. Almost all following papers conclude in the same result when it comes to the shape of the pdf. It is leptokurtic with heavier tails than a Gaussian distribution which is not suitable to model wind power forecast errors. (B.-M. Hodge, Lew, et al. 2012; A. Florita, B. Hodge, and M.R. Milligan 2012; B.-M. Hodge, Anthony Florita, et al. 2012; H Holttinen, Nielsen, and G Giebel 2002; Tewari, Geyer, and Mohan 2011; Fabbri et al. 2005; Pinson and Kariniotakis 2004; Armin Luig, Stefan Bofinger, and H. G. Beyer 2001; Mello, Lu, and Makarov 2011; Lange 2003; Pinson 2006). Interesting correlation analysis is given by Mello, Lu, and Makarov (2011) and H Holttinen, Nielsen, and G Giebel (2002). They discovered that errors have a significant autocorrelation in smaller time lags and are correlated when comparing different forecast horizons. Hour-ahead forecast errors are cross correlated with real-time forecast errors in this study, as well as day-ahead with hour-ahead forecast errors. H Holttinen, Nielsen, and G Giebel (2002) analyze correlation fore smaller forecast horizons, 0-36 hours, and found that the errors cancel out to some extent when combining larger areas/several wind turbines and parks. The same result comes from Fabbri et al. (2005) and Tewari, Geyer, and Mohan (2011). Tewari, Geyer, and Mohan (2011) add that with geographically more dispersed forecasts, the pdf approaches a Gaussian distribution. This corresponds with the fact of a smaller error and a decreasing kurtosis over timescale.

4.2 Economic Forecast Errors

Literature on economic forecast errors focuses on performance evaluation, optimal forecast combination in survey forecasts and comparison of different forecast methods.

Coibion and Gorodnichenko (2008) present three stylized facts about the adjustment of macroeconomic forecast errors after structural shocks. First, forecasts fail to adjust after structural shocks and forecast errors connected to shocks last from about 6 to 12 months. Dovern and Weisser (2011) come to a similar result. Stark et al. (2010) take a closer look at the performance evaluation of the SPF from the Federal Reserve Bank Philadelphia (Philadelphia Fed). Findings are, that the survey's performance falls very sharply at horizons beyond the first quarter, data revisions have a large impact on the survey's performance and that the survey generally outperforms the baseline persistence model. Makridakis and Michele Hibon (2000) give important conclusions from the Makridakis M3 competition. Findings are that complex methods do not necessarily perform better than simpler ones. Also the performance rankings vary with the error measure being used, which is confirmed by Granger and Pesaran (2004), Hyndman and Koehler (2006), Armstrong and Collopy (1992), and Diebold and Lopez (1996). Another result is that various methods combined typically outperform a single method. Those results are confirmed by Makridakis and Michele Hibon (2000), Makridakis, Chatfield, et al. (1993), and Makridakis, Andersen, et al. (1982). Makridakis, Michèle Hibon, et al. (1987) find that there is a larger percentage outside confidence intervals than postulated theroretically, which indicates a heavy-tailed distribution function. Forecast error pdfs lack of normality and possibly exhibit heteroscedasticity, thus volatility clustering. Autocorrelation is found to have little impact except for longer forecast horizons. Rossi and Sekhposyan (2015) and Elliott and Timmermann (2004) analysis correspond to a heavy-tailed pdf. Dovern and Weisser (2011) find that forecast errors are surprisingly high dispersed and pdfs generally exhibit excess kurtosis and are so heavy-tailed. Also they analyze correlation structures but find no common correlations between forecasts.

4.3 Error Correction

There is some literature that addresses the topic error correction or error reduction without using *cointegration and vector error correction model (VECM)* from Engle and Granger (1987) and Johansen (1991). In wind power, for example, Focken et al. (2002) analyze and apply spatial smoothing effects. They find that aggregating power predictions to larger regions reduces the forecast error, compared to single sites. Aggregating larger regions results in higher error reductions. Armin Luig, Stefan Bofinger, and

H. G. Beyer (2001) apply a simple neural network as a postprocessor procedure and archieve a reduction of the initial forecast error.

Engle and Granger (1987) and Johansen (1991) describe the relationship between cointegration and error correction models and provide estimation procedures and tests. Examples are given by A. M. Masih and R. Masih (1997), Mukherjee and Naka (1995) or Bahmani-Oskooee and Alse (1994).

4.4 Summary

As a summary of the literature research findings, common characteristics which can present stylized facts are listed below.

- Wind power and macroeconomic forecast performance rankings vary with the error measure being used.
- Wind power and macroeconomic forecasts get more inaccurate with increasing forecast horizon.
- Wind power and macroeconomic forecast errors typically have a heavy-tailed and leptokurtic pdfs. Kurtosis values decrease with increasing forecast horizon.
- Wind power and macroeconomic forecast errors exhibit significant autocorrelation only in very small timelags.
- A combination of wind power and macroeconomic forecast methods, respectively, usually performs better than single forecasts in that field. Forecast errors of more advanced forecast systems with various methods included are consequently typically smaller.
- Macroeconomic forecast errors exhibit volatility clustering.
- Wind power forecast errors with different forecast horizons of the same variable are correlated.

5 Data Analysis

In this section we present the data sets utilized and conduct the time series analysis. After presenting the data sets, necessary preprocessing steps are explained. In the analysis section, we analyze characteristics of the data sets before estimating a *cointegration* relationship and *vector error correction model*.

5.1 Data

In this section the data and preprocessing are presented. First, we present the data sets with their source and structure. Then necessary preprocessing steps, for example how to handle outliers and missing values, are explained.

5.1.1 Data Sets

For this thesis wind power and marcoeconomic data are utilized. Electrical power transportation in Germany is operated by the four TSOs 50Hertz, TenneT, TransnetBW and Amprion. They publish wind power forecast and actual data in a joined information platform (netztransparenz.de 2017). The forecast is estimated on a daily basis for the following day and typically a combination of several wind energy feed-in forecasts, while actual numbers are a projection of representative wind farms. The data is downloadable in spreadsheet format and lists forecast and real values by the respective TSO in 15 minute periods. With the forecast being estimated at 8 a.m. for TenneT and 9 a.m. for 50Hertz of the respective preceding day, covered forecast horizons from TenneT are so 16 to 40 and from 50Hertz 15 to 39 hours in 15 minute intervals. In this work we use hourly forecast horizons and the respective four 15 minute periods are counted together. For example, the values at 2:00 p.m., 2:15 p.m., 2:30 p.m. and 2:45 p.m. are counted as 2 p.m.. TransnetBW and Amprion do not publish the particular forecast estimation time, therefore we exclude it from our analysis. We use time series of 2012, 2015 and 2016. With four values per forecast horizon per day, the length of the time series for one year is 1460.

Economic data sets in this work are from the Philadelphia Fed SPF. It is a quarterly survey of macroeconomic forecasts in the United States and began in the fourth quarter 1968. It started with five variables and several more have been added over time. The participants come largely from financial firms, banks, consulting firms and research centers (Federal Reserve Bank of Philadelphia 2017c; Croushore 1993). The Philadelphia Fed data is available in spreadsheet format of different categories. For the use of this work the mean responses of the survey participants are analyzed. We use the five variables that have been surveyed since the start of the survey. These time series have 189 data points and we do not use the series with less data points because of their lower significance. The chosen variables have quarterly projections for five quarters and are seasonally adjusted (Federal Reserve Bank of Philadelphia 2017d). The variables are nominal GDP, nominal corporate profits after tax, unemployment rate, average level of the index of industrial production and average level of housing starts.

As mentioned above, forecast errors of shorter forecast horizons are created with the *persistence method* which is defined in section 3.1. The forecast is estimated with actual 50Hertz data from 2015. We estimate forecasts with forecast horizons of 15, 30 and 45 minutes as well as 1, 2, 5 and 10 hours. The forecast is for one year, which gives 35000 data points after preprocessing.

5.1.2 Data Preprocessing

The data sets described in section 5.1.1 are generally of high quality with only few missing values. However, there are a few preprocessing steps necessary.

Calculate forecast error: Since the data is given in actual and forecasted value, the resulting forecast error needs to be calculated as shown in section 3.1.

Split forecasts in different forecast horizons: The time series most often include various forecast horizons. To obtain time series with the particular forecast horizons, they are filtered and separated.

Missing Values: The time series have some missing values, which can be the case due to technical malfunctions in wind power or forecasters who do not submit every survey variable. The average percentage of missing values from 50Hertz and Tennet time series with different forecast horizons is 0,11%. The maximum gap in the data is 8 values, which makes a missing time period of two hours as the values come in 15 minute periods. The average percentage of missing values in Philadelphia Fed's SPF is 0,53%. The maximum gap has four missing values in a row, which makes a time period of one year, as the data is quarterly. Deleting missing values would affect the length of the series which we need to stay the same for correlation analysis. Because of that, the small percentages and the small gaps of missing values, we choose to replace missing values with the average of the value before and after the gap.

Outliers: For the wind power data sets, the respective TSOs state to already exclude unreasonable values which could, for example, result from measurement errors. They are left blank and so they are handled as missing values. However, in a visual examination of the data there are a few potential outliers unveiled. The respective values have been researched and in almost every case they are regular values. Since they are part of the forecast, we do not handle them as outliers. In the macroeconomic data, we find values which seem to be outliers, but come from economic cycles. They are almost always not seen before they occur and so result in large forcast errors in one quarterly survey, while the next survey values are already adjusted to the new conditions. Economic crises will probably occur again and predicting those is one of the tasks of a forecast model, so we assume forecast error fluctuations from crises as regular values.

Besides these steps, there arecommo some other issues which arise in the data preparation process. The first issue is data correction. As previously described in section 4.2, Stark et al. (2010) found that data revisions, which are quiet common in the macroeconomic sector, have a large impact on Philadelphia Fed's SPF's accuracy. Which actual values should be used to calculate the forecast error? This is a challenging question, as the decision could have an impact not just on the performance evaluation, but on other characteristics of the error. First, we look which actual values were used by others. The Philadelphia Fed provides forecast error statistics and the data it is calculated from for every variable used (Federal Reserve Bank of Philadelphia 2017a; Federal Reserve Bank of Philadelphia 2017b). The actual values are revised every quarter. The Philadelphia Fed uses a moving average of those actual values to evaluate the forecast performance. However, analyzing which actual values to use and its impact on the performance and characteristics of the resulting errors goes beyond the scope of this work and will be a topic for further research. We use the first published actual value, that is also closest to the data environment the survey panelists faced at the time the forecast was made. However, it is important to bear that in mind when interpreting the results. Other data sets used in this work have no corrections.

The second issue is normalization. Rescaling data to fixed boundaries is about comparability and in this work it makes sense in the scope of performance evaluation. Armstrong and Collopy (1992) and Madsen et al. (2005) work, for example, with percentage errors to compare forecast performance and Madsen rescales the data with the wind capacity in the scope of wind power prediction performance evaluation. In section 5.2.2 we therefore choose an percentage error measure which is not scale dependent. For other analysis methods we do not rescale the data, as we want to preprocess it as little as possible.

For other analyses in this work, we do not necessarily need rescaling, as we search for common characteristics and our time series analysis.

5.2 Assessment of Time Series Properties

In this section we analyze the time series and search for common characteristics that could present stylized facts. Recalling section 2, stylized facts are simplifying abstractions of statistical properties. Therefore we gather characteristics and properties of our time series and look for commonalities. All figures and calculations are created with Matlab, mainly with its econometric toolbox (Mathworks Documentation 2017a).

5.2.1 Stationarity

In this section we visually examine if the time series are stationary and use hypothesis testing to confirm the results. In figures 1, 13 and 14, the means seems to be quite constant over time. No trend or seasonal effects are apparent. A look at the autocorrelation functions in figure 22 also indicates a stationary process as it drops quickly to zero. In figures 21 and 7 it drops at the beginning but then continues to zero at a slowlier rate within the first 15 lags, which could be a indicator of long-memory stationary processes. (Ruppert 2011a; Brockwell and Davis 2016a). Literature distinguishes between short-memory and long-memory stationary processes. The time indication short or long refers to the mean, if it is constant in short-term or only in long-term. If the autocorrelation function drops to zero exponentially fast, it indicates a short-memory stationary process. If it quickly falls to smaller values and then slowly decays at a polynomial rate well into higher lags, it is a indication of a long-memory stationary process (Ruppert 2011d).

To evaluate stationarity empirically we test the null hypothesis of a non-stationary unit root with an augmented Dickey-Fuller test. The test is defined in B.1 and results are listed in table 2 in the appendix. The rejection

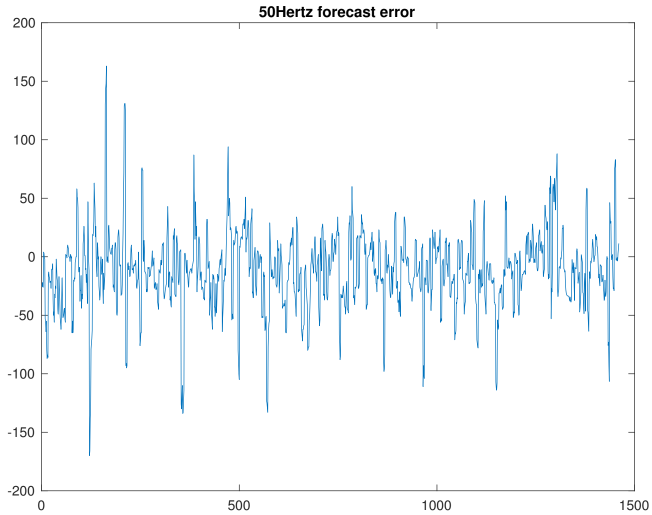

Figure 1: Forecast error of 50Hertz 2016 wind power forecast with a forecast horizon of 15h. Source: Author's illustration of netztransparenz.de 2017 data.

of the null hypothesis, with p-values lower than the smallest tabulated value 0.1, indicates that forecast errors are stationary. However, a Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test gives partially contradictory test results as it rejects the null hypothesis of a stationary process for all 50Hertz and Tennet data as well as for Philadelphia Fed SPF data with forecast horizons higher than one quarter. The respective p-values are clear. The test is defined in B.2 and results are listed in appendix table 3. Such test results indicate a stationary process with long-term memory (Ruppert 2011d; Lee and Schmidt 1996). Figures 13 and 14 are confirmed stationary, and figure 1 is indicated stationary with long-term memory. With the concerning persistence and wind power time series having a size of 35000 and 1460 data points, we consider the test results as reliable. However, the results of Philadelphia Fed SPF data should be considered with caution as they only have 189 data points.

In this section, we conclude that stationarity, either short- or long-memory, is a common characteristic of forecast error time series. Wind power series tend to be stationary processes with long-term memory, whereas macroeconomic series tend to be stationary with short-term memory.

5.2.2 Forecast performance

In this section forecast performance, or accuracy, is analyzed. The goal of this section is not to compare the performance of different methods and models, but rather search for common characteristics. The intuitive assumption is that a forecast gets less accurate and the error increases with increasing forecast horizon. Two error measures are analyzed. In several articles regarding forecast performance evaluation the following two measures are applied in order to balance their respective advantages and disadvanteges (Armstrong and Collopy 1992; Madsen et al. 2005; Makridakis and Michele Hibon 2000). The *RMSE* is defined in equation 12,

$$RMSE_{t+k|t} = \sqrt{\frac{\sum_{t=0}^{N} e_{t+k|t}^2}{N}} \tag{12}$$

where $e_{t+k|t}$ is the error at forecast horizon k (Madsen et al. 2005). One of its disadvantages is scale-dependency, that means comparability is limited when applied to different data sets. The second measure, *Symmetric mean absolute percentage error (sMAPE)*, is defined in equation 13,

$$sMAPE_{t+k|t} = 100 * \sum_{t=0}^{N} \frac{2 * |e_{t+k|t}|}{y_{t+k} + y_{t+k|t}} \tag{13}$$

where y_{t+k} is the actual value at time $t+k$, $y_{t+k|t}$ is the forecast value at time $t+k$ made at time origin t and $e_{t+k|t}$ is the error at forecast horizon k. It is used and partly introduced by Makridakis and Michele Hibon (2000) for evaluating the M-Competition. It is not scale-dependent as it calculates relative errors.

Results for RMSE of the SPF are presented in table 1 and in appendix table 4 for every data set. They show increasing errors over forecast horizon. The

Hor.	CPROF	HOUSING	INPROD	NGDP	UNEMP
1	37.51	0.10	6.81	62.82	0.17
2	58.02	0.16	10.02	101.85	0.38
3	73.23	0.21	12.68	139.26	0.57
4	87.47	0.25	14.99	176.73	0.76
5	99.37	0.30	17.09	213.92	0.94

Table 1: RMSE of various Philadelphia Fed SPF forecasts for increasing forecast horizon. The RMSE increases significantly over forecast horizon. Source: Author's illustration of Philadelphia Fed SPF data.

same characteristic can be seen in appendix tables 5, which shows $sMAPE$ for the persistence forecast, and 6, which shows $sMAPE$ for 50Hertz data. There are differences between the forecast methods and the forecast horizon intervals apparent. The Philadelphia Fed SPF uses mean values of very advanced forecasts and has large forecast horizon intervals of one quarter. The error increases significantly here. The wind power forecasts of German TSOs are very advanced, but with shorter forecast horizon intervals of only a few hours, In table 6 the error only slightly increases. The persistence forecast in table 5 increase, despite of even smaller forecast horizon intervals, more significantly because of the simple forecast method. Please note that the RMSE values should only be directly compared with caution as the underlying forecasts are of different scale.

The common characteristic identified in this section is that forecast errors generally increase over forecast horizon. The increase is more significant for the low-frequency macroeconomic data than for the high-frequency wind power data, which increases only slightly over forecast horizon. This is probably caused by higher forecast horizon intervals and more available data points for high-frequency data.

5.2.3 Conditional density functions

In the following section properties of conditional forecast error distributions are examined. The shape and behaviour for increasing forecast horizon is of special interest to us, as the increasing error measures from section 5.2.2 should have an impact on conditional error pdfs. We also want to test if the forecast errors follow a Gaussian distribution.

To examine this histograms and kernel density estimations of our data are created and analyzed. In addition to the first and second statistical moments mean and variance, the third and fourth, skewness and kurtosis, are analyzed. Table 7 shows the four statistical moment values for all our datasets.

The third statistical moment skewness is a measure of asymmetry around the sample mean. It is defined in equation 14.

$$s = \frac{E(y-\mu)^3}{\sigma^3} \tag{14}$$

where y is a distribution, μ the mean and σ the standard deviation of y. Symmetric distributions have a skewness of 0. A negative (positive) skewness indicates a left (right) skew of the data compared to the mean (Ruppert 2011e). Figure 4 and 19 show a visualization of typical skewness values. The analysis can not show characteristic skewness values or trends. However, the visual analysis suggests a slight trend to symmetry with increasing forecast horizon. But as this trend is only partially true in our data, typical characteristics in skewness values are not apparent.

The fourth statistical moment kurtosis indicates the tailedness and peakedness of a distribution compared to the normal distribution and is defined in equation 15.

$$k = \frac{E(y-\mu)^4}{\sigma^4} \tag{15}$$

where y is a distribution, μ the mean and σ the standard deviation of y. A normal distribution has a kurtosis of 3. *Excesskurtosis*, defined as $k-3$,

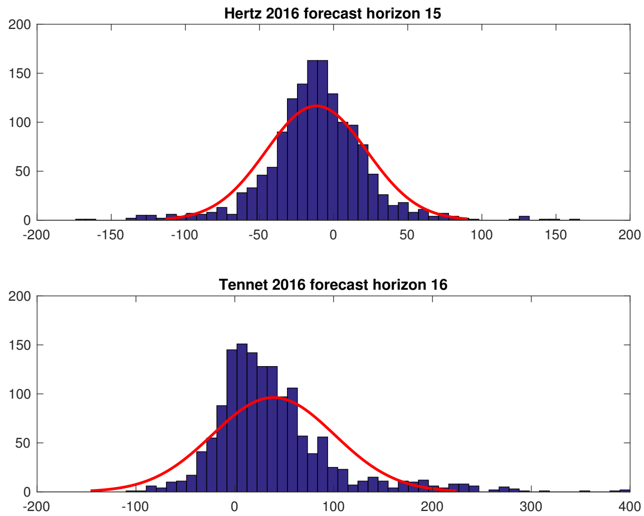

Figure 2: Histograms of 50Hertz and Tennet wind power forecast errors with a Gaussian overlay of the same mean and variance. The forecast horizon is 15 hours for 50Hertz and 16 hours for Tennet. The 50Hertz data has a mean of -11.59, a variance of 1152.95, a skewness of -0.04 and a kurtosis of 6.38. The Tennet data has a mean of 39.11, a variance of 3806.24, a skewness of 1.73 and a kurtosis of 7.76. A visual examination of the conditional pdf's shows more pronounced peak, steeper shoulders and heavier tails compared to the respective Gaussian distribution. Source: Author's illustration of netztransparenz.de 2017 data.

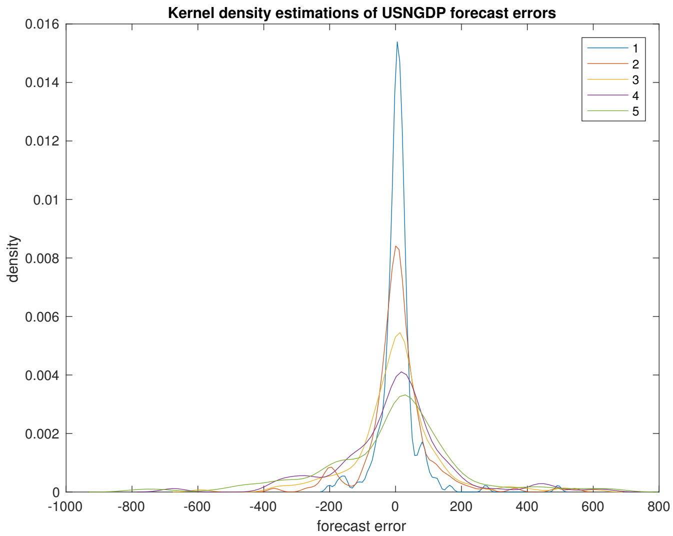

Figure 3: Kernel density estimation of US NGDP with forecast horizons one to five quarters. With increasing forecast horizon, the conditional pdf gets less pronounced in its peak, the shoulders get flatter and the tails get heavier. Source: Author's illustration of Philadelphia Fed SPF data.

compares kurtosis values to those of the normal distribution (DeCarlo 1997). Balanda and MacGillivray (1988) critically review the interpretation of the kurtosis. They define that kurtosis measures the movement of probability mass from the shoulders of a pdf to its center and tails. A positive *excess kurtosis* indicates heavier tails and a higher peak, often referred to as *leptokurtic*. A negative *excess kurtosis* indicates a flatter distribution with lighter tails, often referred to as *platykurtic*. Figure 5 and 20 show the kurtosis values for our data sets. There is a clear trend of decreasing kurtosis values over increasing forecast horizon, which is confirmed for almost all our datasets in table 7.

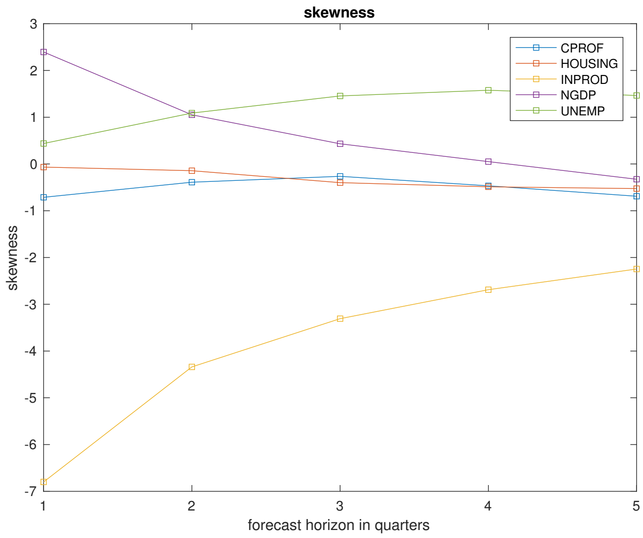

Figure 4: Skewness of various Philadelphia Fed SPF variables with forecast horizons one to four quarters. The skewness here has a slight trend to symmetry with increasing forecast horizon. Source: Author's illustration of Philadelphia Fed SPF data.

Now we take a look at the histograms, which are shown in figures 2, 17 and 18. Shared characteristics of the visual examinations are that all conditional forecast error pdfs analyzed have a more pronounced peak, steeper shoulders and fatter or heavier tails than the corresponding Gaussian pdfs. These

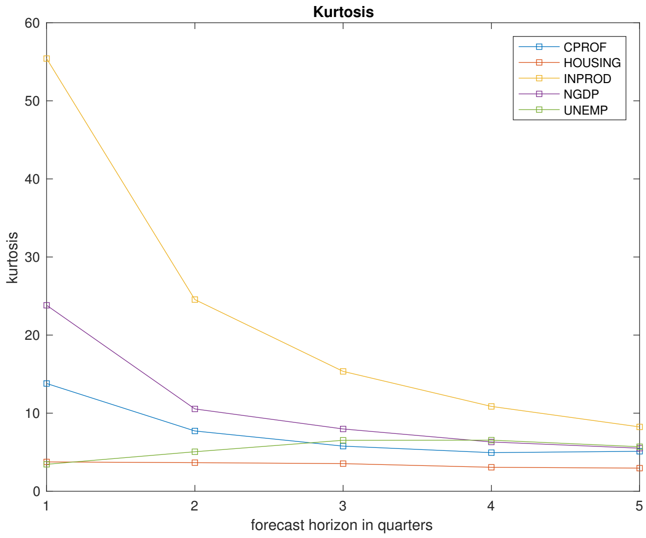

Figure 5: Kurtosis of various Philadelphia Fed SPF variables with forecast horizons one to four quarters. The values reach from 3.45 to 55.41. The kurtosis here has a clear decreasing trend with increasing forecast horizon.Source: Author's illustration of Philadelphia Fed SPF data.

characteristics are consistent with the positive excess kurtosis values of the underlying data. That means that the majority of the errors are close around the mean with a few very large errors, which build the heavy tails of the conditional forecast error pdf. To compare the conditional pdfs of different forecast horizons we use kernel density estimation, which estimates pdfs more precisely than histograms (Ruppert 2011f). Figures 3, 15 and 16 show us kernel density estimations for increasing forecast horizon. The changes in the conditional forecast error pdf appear very clearly. The peak gets less pronounced and the probability mass shifts to the shoulders. This is in accordance to the decreasing kurtosis values from figures 5 and 20, however in the visual examinations the tails seem to get heavier. The interpretation for that probability mass shift is that for short forecast horizons, a majority of errors are in the center, which means close around the mean, with few larger errors. With increasing forecast horizon, the error range expands, large errors in the shoulders and very large errors in the tails occur more often. Except the heavier tails, the shape of the conditional forecast error pdfs approaches a Gaussian distribution with increasing forecast horizon. Excess kurtosis values and the shape of the conditional pdfs indicate a non-Gaussian pdf, which can be clearly seen in a visual examination of figures 2, 17 and 18. These indications are confirmed by a *Kolgomorov smirnoff tests*. The null hypothesis of normal distribution is rejected in every case with p-values below E-8, except for US industrial production index with a forecast horizon 1h. In this case the p-value is 0.1431, so the result is not very clear. The test results are presented in table 8 in the appendix B.3.

Common characteristics of conditional forecast error pdfs, are *heavy-tails* and *excesskurtosis*. Compared to corresponding Gaussian pdfs, they exhibit more pronounced peaks, slighter shoulders and heavier tails. If the forecast horizon increases, the heavy-tailed and excesskurtosis feature decreases. The peak gets flatter and the shoulders get fatter.

5.2.4 Correlation analysis

In this section we examine the autocorrelation and correlation structure of our data sets. First, we take a look at the autocorrelation and then crosscorrelation structures. Equation 16 shows the sample autocorrelation function, where $\hat{\gamma}$ is the covariance and h the lag. Broadly speaking, it measures the similarity, or correlation, of a series with itself as a function of lag h (Brockwell and Davis 2016b).

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < -n \tag{16}$$

Figures 7, 21 and 22 show typical autocorrelation functions or our data sets, which have already been described in section 5.2.1. Partial autocorrelation functions are included for completeness. The assumption is, that there should not be high and persisting autocorrelation values in forecast errors.

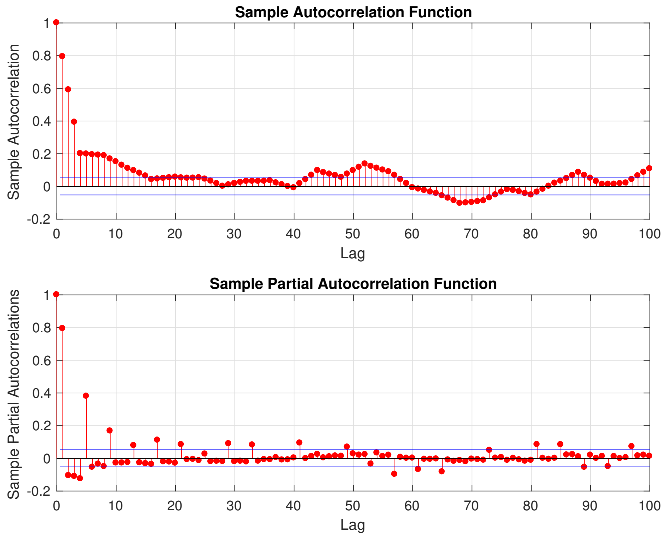

Figure 6: Autocorrelation function and partial autocorrelation function of Tennet2015 data with a forecast horizon of 39 hours. Source: Author's illustration of netztransparenz.de 2017 data.

They could be anticipated in the forecast model, for example if there was a trend in the forecast error which is indicated by high autocorrelation values. Cont (2005) recalls the autocorrelation behaviour of stationary processes. A stationary process has *long-range dependence*, if the autocorrelation function decays slowly at a power law rate. It has *short-range dependence*, if it decays at a geometric, or exponential, rate. Cont mentions that there are also other, intermediate decay rates possible and so the interpretation is not always clear. Most of our autocorrelation functions drop quickly to insignificant values, indicating short range dependence. There are also a few which decay at a slowlier rate, indicating more long-range dependence, like in figure figure 6. Here the decay goes into higher lags about 25. The rate of decay varies with the sector the forecast comes from. For our US macroeconomic data the autocorrelation values are only significant in the first few lags, until about 5. The decay typically takes longer in wind power data, from around 10 to roughly 30 lags. Another common characteristic is that the autocorrelation sporadically jumps to small, significant values in higher lags, for which figure 6 gives an example. In the wind power data the periods of small, significant autocorrelation in higher lags are longer than in the US macroeconomic data. Also a property for all variables is that the autocorrelation generally increases with increasing forecast horizon.

A quantitative criteria of existing significant autocorrelations is given by Ljung-Box Q test results. It tests the null hypothesis of no significant autocorrelation, depending on the lag. The test is defined in appendix B.4 and table 9 shows the results. The null hypothesis is rejected in 52 out of 56 cases. The variables where the null hypothesis can not be rejected are US macroeconomic variables with forecast horizon 1h. This is a sign of autoregressive structure in most of our data.

For the interpretation of the autocorrelation one should consider its reliability. Cont (2001) states, that a heavy-tailed feature of time series can make its interpretation problematic. He summarizes that such autocorrelation functions can be highly unreliable and may have large confidence intervals. We have to keep that in mind for the autocorrelation results above.

We calculated Pearson correlation coefficients between different forecast horizons of the same forecasts, to examine underlying correlation structures. Equation 17 shows the Pearson correlation coefficient, where $cov(x, y)$ is the covariance of two time series x and y, and σ is the standard deviation. Broadly speaking, it measures the similarity, or linear correlation, of two series. It ranges from -1 to 1. -1 is called *perfect negative linear correlation*, 0 *no linear correlation* and 1 *perfect positive linear correlation*

$$\rho(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y} \tag{17}$$

In tables 10, 11 and 12, it can be seen that forecasts with different forecast horizons of the same variable are correlated. The correlation coefficients are highest for the respective next lower and higher horizon and then decrease.

The results of this section are that forecast errors typically have an autore-

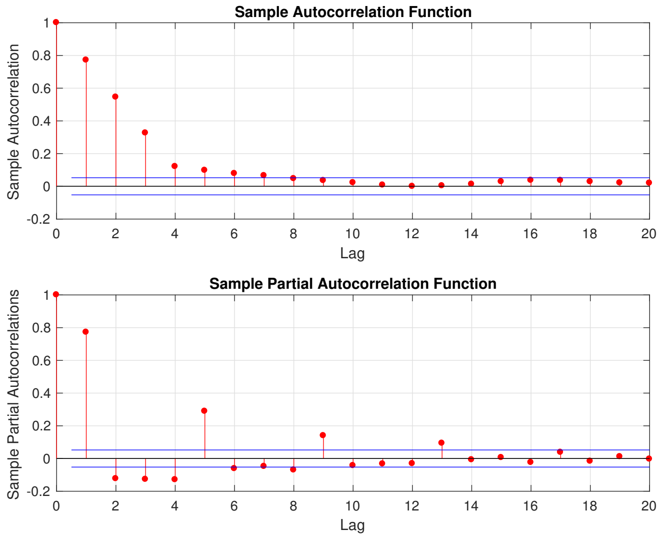

Figure 7: Autocorrelation function and partial autocorrelation function of Hertz 2016 data with a forecast horizon of 15 hours. Source: Author's illustration of netztransparenz.de 2017 data.

gressive structure. The values of the autocorrelation functions drop quickly to smaller values. The Philadelphia Fed SPF data only shows significant autocorrelation in the first few lags, indicating short-term dependence. The wind power data shows long-term dependence, with slowlier decaying autocorrelation values. Another finding is that forecasts with different forecast horizons of the same variable are correlated.

A detailed analysis of possible correlation relationships within single sectors, or forecasts for the same period, goes beyond the scope of this work and it will be a challenge to gain extensive data sets. Especially correlations between forecasts for the same topic and period are of special interest. However, this could reveal interesting relationships like reaction to economic shocks or extreme weather situations. Also relationships with the actual and forecast time series could be interesting, for example if there are properties which drive volatility in the forecast error time series. These examples will be topic for further research.

5.2.5 Volatility clustering

In this section we examine if forecast errors exhibit volatility clustering, such called autoregressive conditional heteroscedasticity (ARCH) effects. Volatility measures the fluctuation of a time series using the *standard deviation* (Auer and Rottmann 2011). *Volatility clustering* is a well known stylized fact in financial time series and describes the property that large variations tend to be followed by large variations, just like small variations by small variations. Thus large and small fluctuations cluster together (Cont 2001; Cont 2007). ARCH effects can be modeled and are a special case of such called generalized autoregressive conditional heteroscedasticity (GARCH) models. A $GARCH(p,q)$ model is shown in equation 18 and 19, where a series a_t consists of the standard deviation σ_t and a white noise error term ϵ_t.

$$a_t = \sigma_t \epsilon_t \tag{18}$$

Equation 19 shows the relating standard deviaton, which consists of a sum of the last p weighted values a_{t-1} and a sum of the last q weighted standard deviations σ_{t-1}. The model works as follows: If the last p values and so the last q standard deviation tends to be high (small), the process a_t tends to be high (small), too. Thus high and standard deviations cluster togehter, respectively. (Ruppert 2011g; Engle 1982)

$$\sigma_t = \sqrt{\omega + \sum_{i=1}^{p} \alpha_i a_{t-1}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-1}^2} \tag{19}$$

A first indication of this phenomenon can be given by a series itself and a moving average volatility plot. Separated periods of large and small fluctuations indicate *volatility clustering.* Figures 1, 13, 14 and 8 show our example time series and their respective moving average volatility plots. The 50Hertz series indicates high volatility at the beginning with ongoing smaller cycles of volatility. In the persistence series the volatility clusters are more clearly, while the US unemployment series has high volatility periods at the beginning and at the end with small volatility in the middle section. However, one should keep in mind that the US unemployment data has very little datapoints and is so not as reliable as the first two examples.

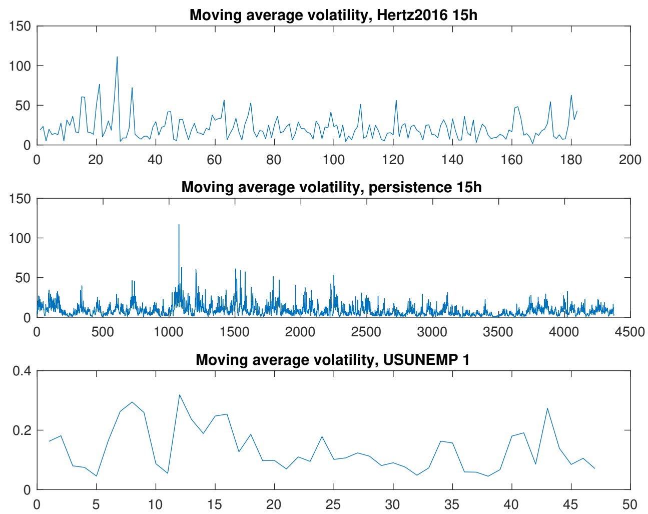

Figure 8: Moving average volatility of 50Hertz 2016 15h, Persistence 15min and US unemployment rate one quarter ahead. The x-axis shows the timescale in tics. The respective moving average intervals are 8, 8 and 4. They have been choosen in terms of visual clarity. Source: Author's illustration of netztransparenz.de 2017 data.

Another expression of *volatility clustering* is that squared time series exhibit significant autocorrelation. Figure 9 shows the autocorrelation functions of our squared example errors. Both 50Hertz and the Persistence squared errors show significant autocorrelation. At the top subplot, the values are only significant until lag 6. In the middle subplot, autocorrelation values stay significant well into higher lags. At the bottom figure, most of the values from lag 1 on are not significant. We refer to section 5.2.4 to keep the autocorrelation function's reliability with heavy-tailed time series in mind.

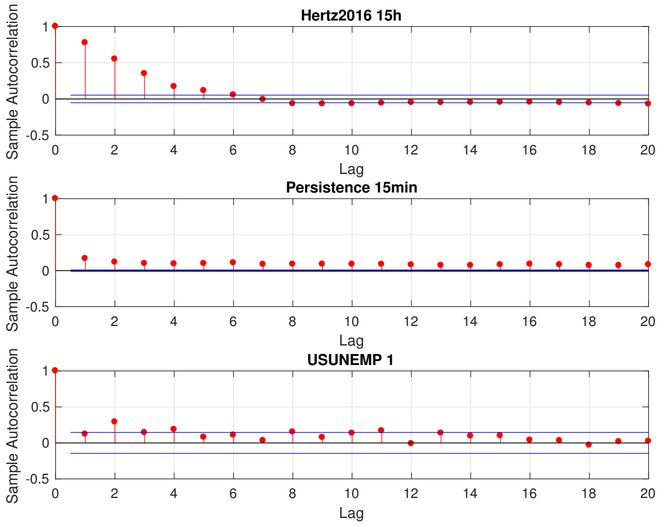

Figure 9: Autocorrelation of 50Hertz 2016 squared forecast errors with a forecast horizon of 15 hours. Source: Author's illustration of netz-transparenz.de 2017 data.

In order to confirm our results we conduct *Engle's ARCH test* on all our data sets to assess whether forecast error time series generally exhibit volatility clustering. Therefore it tests whether a squared series exhibits significant autocorrelation (Engle 1982). The test is described in Appendix B.5 and the results are listed in table 13. The null hypothesis of no *ARCH effects* is rejected in 52 out of 56 cases with very small p-values ranging from E-6 to zero. This confirms the visual results and indicates that our forecast error series generally exhibits *volatility clustering*. This phenomenon could partly be explained in the context of forecast errors with actual time series exhibiting calm and unsteady periods. For example extreme weather situations or stormy periods and economic shocks, which can often not be accurately forecasted, thus driving volatility in the forecast error. After such shocks, the forecast needs time to adjust. This could be topic for further research and also leads to the next section, *error correction*. The result of the analysis in this section is, that forecast error series generally exhibit volatility clustering. This is an important finding as once large or small errors occur, they tend to stay large, or small, respectively. Thus once a forecast value is inaccurate, it probably remains inaccurate over the next several periods. The same behaviour applies to accurate forecast values. This finding can be implemented in risk management approaches and it will be topic for further research how to exactly do that.

5.2.6 Cointegration and Error Correction

As previously mentiond in section 3.4, two or more nonstationary time series are said to be co-integrated, if they have a similar behaviour and trend. We can use this relationship to find a linear combination which is detrended and thus stationary. After finding cointegrated time series, the next step is to estimate an *error correction model*, which models the linear combination's short-term deviation from its mean, the *error*, and corrects it. The general references used for the model estimation in this section are Mathworks Documentation (2017e), Johansen (1991), and Kilian and Lütkepohl (2016).

We will first visually examine potential cointegration relationships and then apply the *Johansen cointegration test*, which tests for cointegrating relationships and estimates the coefficients of a VECM. As multivariate time series we use the variables from Philadelphia Fed SPF. They turn out to to track each other fairly closely within a forecast of different forecasts horizon, which can be seen in figure 10. They are non-stationary because of the underlying trend in the data. We choose the variable USCPROF, the nominal corporate profits after tax of the United States, as example in this section. The similar trend and behaviour in figure 10 indicates a cointegration relationship between the respective variables.

In the Johansen (1991) approach, testing for cointegration and parameter estimation of a *VECM*, are combined in one step. The test assesses the rank of Π, which is the number of cointegration vectors in a VECM and shown

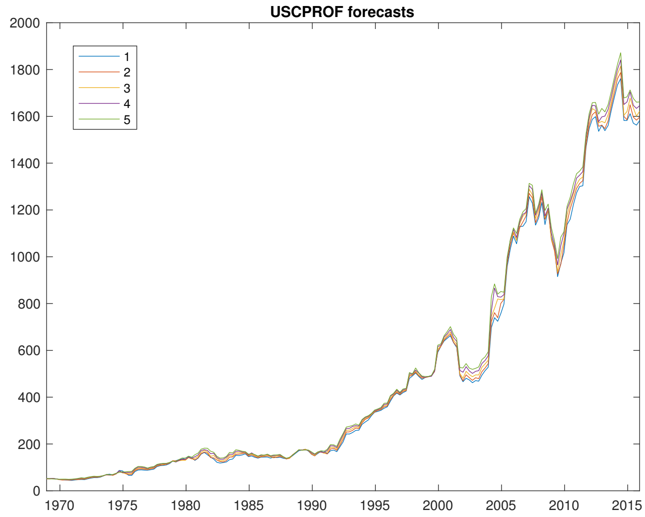

Figure 10: Forecast error plot of US nominal GDP with forecast horizon 1 to 5 quarters. Source: Author's illustration of Philadelphia Fed SPF data.

in equation 11 and 22. The hypotheses are shown in equation 20, where the null hypothesis means that the cointegration rank is the current tested rank, while a rejection means that there are more cointegration relationships than currently tested. K is the number of time series.

$$H_0(r_0) : rank(\Pi) = r_0, \quad H_1(r_0) : rank(\Pi) > r_0; \quad r_0 = 0, 1, ..., K-1 \quad (20)$$

The test is a sequential procedure, it iterates through $r_0 = 0, 1, ..., K-1$, and stops at the first r_0 for which the null hypothesis can not be rejected. This rank r_0 is the number of cointegration relationships. There are two special cases, the first is $r_0 = 0$, which means that there no cointegrating relationships. The second is $r_0 = K - 1$, which means that the variables are already stationary, thus a VECM makes no sense. Equation 21 shows the test statistic, where $l(r_0)$ is the maximum of the Gaussian likelihood function and λ_i is the i^{th} eigenvalue obtained by the coefficient estimation.

$$t_{jci,trace}(r_0) = 2(\log l(r_0) - \log l(K)) = -T \sum_{i=r_0+1}^{K} \log(1 - \lambda_i) \tag{21}$$

For the maximum likelihood estimation of coefficients and estimation of critical values we refer to Mathworks Documentation (2017e), Johansen (1991), and Kilian and Lütkepohl (2016), where also the definitions and notations of this section come from.

The test is also described in appendix B.6 and conducted with the implemented *jcitest* in matlab (Mathworks Documentation 2017e). We will fit a VECM(1), which is shown in equation 22, where $\Pi = \alpha\beta^T$, $\Gamma_1 \Delta Y_{t-1}$ are weighted lagged differences, μ is a mean vector and D_t a vector of seasonal dummies.

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \Phi D_t + \epsilon_t \tag{22}$$

The form of the error correction term of the estimated Matlab default model, which eliminates stochastic and deterministic trends in the data, is shown in equation 23.

$$\Pi Y_{t-1} + \Phi D_t = \alpha(\beta' Y_{t-1} + c_0) + c_1 \tag{23}$$

The calculated stationary cointegration relationship with cointegration weights β is shown in equation 24, plotted in figure 11 and seems stationary as expected.

$$\beta' Y_{t-1} + c_0 \tag{24}$$

The estimated VECM is shown in equation 25 and plotted in figure 11.

$$\begin{aligned}\Delta Y_t = &\begin{pmatrix} 9.6185 \\ 6.1861 \end{pmatrix} \left(\begin{pmatrix} -0.1828 \\ 0.1811 \end{pmatrix}^{\mathsf{T}} Y_{t-1} + \begin{pmatrix} 0.5133 \\ ... \\ 0.5133 \end{pmatrix}^{\mathsf{T}} \right) \\ &+ \begin{pmatrix} -0.9889 & 1.5376 \\ ... & ... \\ -0.9889 & 1.5376 \end{pmatrix}^{\mathsf{T}} + \begin{pmatrix} 0.4152 & -0.2865 \\ 0.6008 & -0.4124 \end{pmatrix} \Delta Y_{t-1} + \epsilon_t\end{aligned} \tag{25}$$

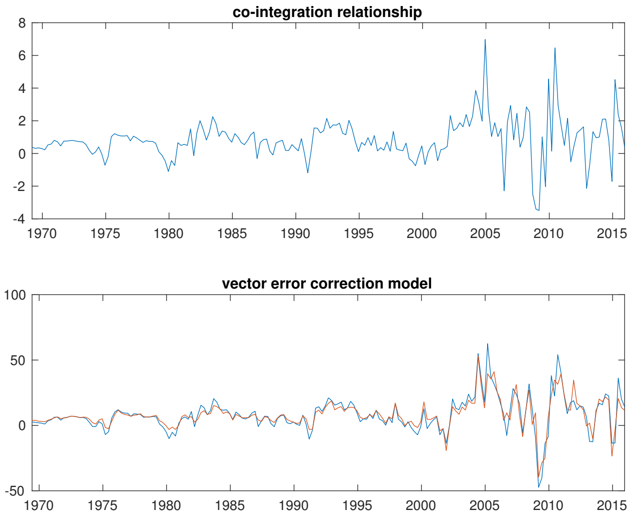

Figure 11: Author's illustration of Philadelphia Fed SPF data. The upper figure shows the cointegration relationship. The lower figure shows the estimated VECM. The first figure is inspired by Mathworks Documentation (2017e)

Every VECM has its equivalent *vector autoregressive (VAR) model*, which estimation is shown in equation 26 and plotted in figure 12. While a VECM models the differenced time series, its equivalent VAR model models the values. In a *VAR(p) model*, values are estimated with a regression on the last p lagged values and an error term. A *VECM(p)* model can be transformed to a *VAR(p+1)* model. In our vase, the *VECM(1)* model is transformed to a *VAR(2)*. We refer to Mathworks Documentation (2017b) for a extended definition of VAR models and its implemented Matlab function.

$$Y_t = \begin{pmatrix} -0.9889 & 1.5376 \\ ... & ... \\ -0.9889 & 1.5376 \end{pmatrix}^{\mathsf{T}} + \begin{pmatrix} -0.3434 & 1.4550 \\ -0.5302 & 1.7077 \end{pmatrix} Y_{t-1}$$
$$+ \begin{pmatrix} -0.4152 & 0.2865 \\ -0.6008 & 0.4124 \end{pmatrix} Y_{t-2} + \epsilon_t \quad (26)$$

As figure 12 shows, the VECM estimation is very accurate. The two VECM estimatations seem to anticipate the actual forecast values a little in advance. This presents the functionality of an *error correction model*, as it pushes short-term deviations "back" to its long-term mean. However, the RMSEs of the SPF forecasts are 14.91 for forecast horizon one quarter and 39.19 for two, while the VECM RMSEs are 37.80 and 55.93. A reason for the higher RMSE values can be a inaccurate trend assessment or too little used lags in the model. For a clear presentation of the basic concept, the VECM estimates a linear trend in the data, while the actual underlying trend in figure 10 seems more quadratic. Johansen provides several extended methods to model different trends in the data (Mathworks Documentation 2017e; Johansen 1991).

The goal of this section was to demonstrate the concept of *error correction* and its potential application to forecast error reduction. A VECM has been estimated and compared with the original forecasts. However, an error reduction could not be shown based on RMSEs. It will be topic for further research to examine the use of cointegration and error correction in the scope of forecast error reduction.

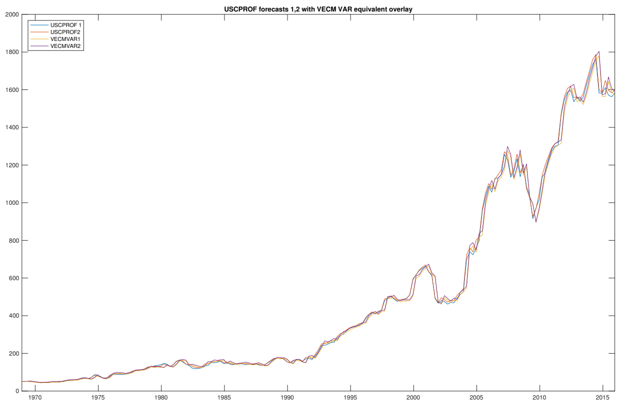

Figure 12: Forecast error plot of US nominal GDP with forecast horizon 1 and 2 quarters with overlay of the respective VAR equivalent. Source: Author's illustration of Philadelphia Fed SPF data.

6 Conclusion

At the end of this work we have to ask ourselves if and how stylized facts can be used to reduce forecast errors. We think that knowledge on forecast errors can indeed be helpful for the improvement and useful in the interpration of the forecasts. For example, with the common presence of *volatility clustering*, we can to some extent expect a period of higher forecast errors once some appear. With the knowledge, that conditional forecast error pdfs are typically heavy-tailed, large forecast errors can be expected.

The main objective of this work was to find common characteristics of forecast errors. Relevant literature has been researched, and the main findings are increasing forecast errors over forecast horizon, heavy-tailed and leptokurtic conditional error pdfs, and significant autocorrelation only in small timelags. Time series characteristics of various wind power and economic forecast errors have been examined. The author's key contribution is a profound examination of time series characteristics and the comparison of forecast error from different sectors. Joint characteristics of forecast errors from different sectors have rarely been presented in previous literature.

As an outcome of analysis and literature research we found the following common characteristics which we propose as stylized facts.

1. Stationarity: Forecast errors are typically stationary, some with indications of long-memory. Wind power series tend to be stationary processes with long-term memory, which means that the mean is constant only in long-term, not in short-term. Macroeconomic series tend to be stationary with short-term memory.

2. Forecast Performance: Forecast error measures typically increase over forecast horizon.

3. Heavy tails: Conditional pdfs of forecast errors are typically not Gaussian, but heavy tailed and leptokurtic. The heavy-tailedness decreases over forecast horizon with decreasing kurtosis values, representing generally larger errors.

4. Correlation analysis: Forecast errors typically exhibit significant autocorrelation. The values of the autocorrelation functions drop quickly to smaller values. The Philadelphia Fed SPF data only shows significant autocorrelation in the first few lags, indicating short-term dependence. The wind power data shows long-term dependence, with slowlier decaying autocorrelation values. Another finding is that forecasts of the same variable with different forecast horizon are correlated.

5. Volatility clustering: Most squared forecast errors show significant autocorrelation values, which indicates volatility clustering. That means that periods of large and small fluctuations tend to cluster together.

It is important, that the findings of this work are verfied in further research. This includes, if hypothesis test results and autocorrelation functions are biased by heavy-tailed distributions, which was mentioned by Cont (2001).

More topics for further research include the impact of data corrections, the analysis of forecast combinations and correlation analysis of actual or forecast series with the forecast error. There might be interesting relationships which drive the volatility of forecast errors.

Additionally, we presented the concept of *cointegration and error correction* and successfully estimated a cointegration relationship, as well as a VECM. The application of VECMs in the scope of forecast error reduction is topic for further research. We see potential in that field, especially to adjust structural shocks. We believe that VECMs can be helpful to optimize the forecast in rough periods.

7 Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt und alle benutzten Hilfsmittel und Quellen vollständig angegeben zu haben, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht zu haben und die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis beachtet zu haben.

Datum *Name*

Appendices

A Additional Analysis Results

Variable	Result	p-value	Variable	Result	p-value	Variable	Result	p-value
Hertz2012 15h	1	0.001	Tennet2012 16h	1	0.001	USHOUSING 3h	1	0.001
Hertz2012 20h	1	0.001	Tennet2012 21h	1	0.001	USHOUSING 4h	1	0.001
Hertz2012 30h	1	0.001	Tennet2012 31h	1	0.001	USHOUSING 5h	1	0.001
Hertz2012 38h	1	0.001	Tennet2012 39h	1	0.001	USINDPROD 1h	1	0.001
Hertz2015 15h	1	0.001	Tennet2015 16h	1	0.001	USINDPROD 2h	1	0.001
Hertz2015 20h	1	0.001	Tennet2015 21h	1	0.001	USINDPROD 3h	1	0.001
Hertz2015 30h	1	0.001	Tennet2015 31h	1	0.001	USINDPROD 4h	1	0.001
Hertz2015 38h	1	0.001	Tennet2015 39h	1	0.001	USINDPROD 5h	1	0.001
Hertz2016 15h	1	0.001	Tennet2016 16h	1	0.001	USNGDP 1h	1	0.001
Hertz2016 20h	1	0.001	Tennet2016 21h	1	0.001	USNGDP 2h	1	0.001
Hertz2016 30h	1	0.001	Tennet2016 31h	1	0.001	USNGDP 3h	1	0.001
Hertz2016 38h	1	0.001	Tennet2016 39h	1	0.001	USNGDP 4h	1	0.001
Persistence 1h	1	0.001	USCPROF 1h	1	0.001	USNGDP 5h	1	0.001
Persistence 10h	1	0.001	USCPROF 2h	1	0.001	USUNEMP 1h	1	0.001
Persistence 15min	1	0.001	USCPROF 3h	1	0.001	USUNEMP 2h	1	0.001
Persistence 2h	1	0.001	USCPROF 4h	1	0.001	USUNEMP 3h	1	0.001
Persistence 30min	1	0.001	USCPROF 5h	1	0.001	USUNEMP 4h	1	0.001
Persistence 45min	1	0.001	USHOUSING 1h	1	0.001	USUNEMP 5h	1	0.001
Persistence 45min	1	0.001	USHOUSING 2h	1	0.001			

Table 2: Augmented Dickey-Fuller test results with p-values for all data sets. The null hypothesis of a unit root process is rejected for significance level alpha = 0.05 with the lowest possible p-values. These results implicate that all analyzed forecast error time series are stationary. Source: Author's illustration with data from netztransparenz.de 2017 and Philadelphia Fed SPF.

Variable	Result	p-value	Variable	Result	p-value	Variable	Result	p-value
Hertz2012 15h	1	0.010	Tennet2012 16h	1	0.010	USHOUSING 3	1	0.010
Hertz2012 20h	1	0.015	Tennet2012 21h	1	0.010	USHOUSING 4	1	0.010
Hertz2012 30h	1	0.010	Tennet2012 31h	1	0.010	USHOUSING 5	1	0.010
Hertz2012 38h	1	0.010	Tennet2012 39h	1	0.010	USINDPROD 1	0	0.10
Hertz2015 15h	1	0.010	Tennet2015 16h	1	0.010	USINDPROD 2	0	0.10
Hertz2015 20h	1	0.010	Tennet2015 21h	1	0.010	USINDPROD 3	0	0.10
Hertz2015 30h	1	0.010	Tennet2015 31h	1	0.010	USINDPROD 4	0	0.068
Hertz2015 38h	1	0.010	Tennet2015 39h	1	0.010	USINDPROD 5	1	0.023
Hertz2016 15h	1	0.010	Tennet2016 16h	1	0.010	USNGDP 1	0	0.10
Hertz2016 20h	1	0.010	Tennet2016 21h	1	0.010	USNGDP 2	1	0.027
Hertz2016 30h	1	0.010	Tennet2016 31h	1	0.010	USNGDP 3	1	0.010
Hertz2016 38h	1	0.010	Tennet2016 39h	1	0.010	USNGDP 4	1	0.010
P15	0	0.10	USCPROF 1	0	0.10	USNGDP 5	1	0.010
P30	0	0.10	USCPROF 2	1	0.043	USUNEMP 1	0	0.10
P45	0	0.10	USCPROF 3	1	0.010	USUNEMP 2	0	0.10
P15	0	0.10	USCPROF 4	1	0.010	USUNEMP 3	1	0.027
P2	0	0.10	USCPROF 5	1	0.010	USUNEMP 4	1	0.010
P5	0	0.10	USHOUSING 1	0	0.069	USUNEMP 5	1	0.010
P10	0	0.10	USHOUSING 2	1	0.0147			

Table 3: Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test results with p-values for all data sets. Source: Author's illustration with data from netztransparenz.de 2017 and Philadelphia Fed SPF.

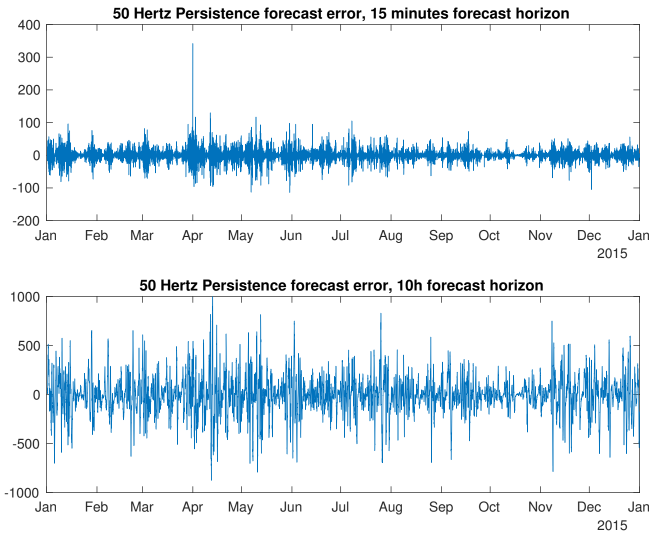

Figure 13: Forecast error of 50Hertz 2015 wind power persistence forecast with a forecast horizon of 15 minutes and 10h. in the upper figure, the excessive value around April is a regular value. It can be clearly seen the forecast error with 10 hours forecast horizon is larger than with 15 minutes. Also, the time series looks like a stationary process with no trend. Source: Author's illustration of netztransparenz.de 2017 data.

Variable	RMSE	Variable	RMSE	Variable	RMSE
Hertz2012 15	181.77	Hertz2015 15	66.35	Hertz2016 15	35.87
Hertz2012 20	201.16	Hertz2015 20	58.55	Hertz2016 20	42.31
Hertz2012 30	220.98	Hertz2015 30	72.81	Hertz2016 30	50.46
Hertz2012 38	265.42	Hertz2015 38	78.88	Hertz2016 38	50.19
Tennet2012 16	188.97	Tennet2015 16	116.56	Tennet2016 16	73.03
Tennet2012 21	172.40	Tennet2015 21	120.03	Tennet2016 21	72.86
Tennet2012 31	223.70	Tennet2015 31	150.29	Tennet2016 31	95.93
Tennet2012 39	213.78	Tennet2015 39	156.72	Tennet2016 39	92.21
USHOUSING 1	0.10	USINDPROD 1	6.81	USNGDP 1	62.82
USHOUSING 2	0.16	USINDPROD 2	10.02	USNGDP 2	101.85
USHOUSING 3	0.21	USINDPROD 3	12.68	USNGDP 3	139.26
USHOUSING 4	0.25	USINDPROD 4	14.99	USNGDP 4	176.73
USHOUSING 5	0.30	USINDPROD 5	17.09	USNGDP 5	213.92
USUNEMP 1	0.17	USCPROF 1	37.51	P15	13.78
USUNEMP 2	0.38	USCPROF 2	58.02	P30	23.44
USUNEMP 3	0.57	USCPROF 3	73.23	P45	31.66
USUNEMP 4	0.76	USCPROF 4	87.47	P1	39.42
USUNEMP 5	0.95	USCPROF 5	99.37	P2	67.72
				P5	130.67
				P10	192.50

Table 4: RMSE of all analyzed forecasts. The RMSE increases significantly over forecast horizon. Source: Author's illustration of netztransparenz.de (2017) and Philadelphia Fed SPF data.

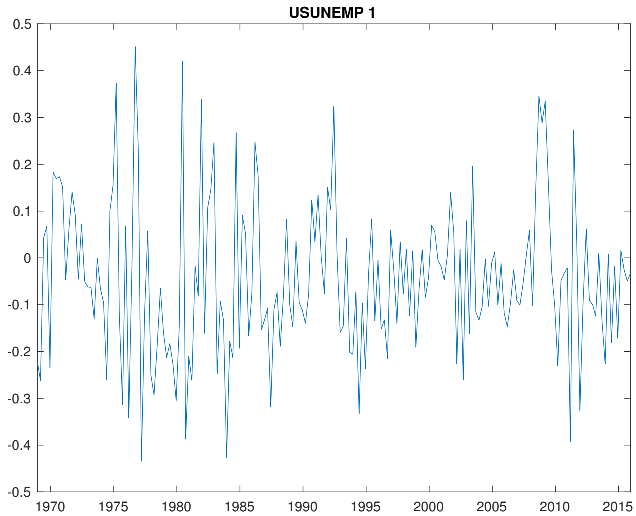

Figure 14: Forecast error plot of US unemployment rate with forecast horizon 1 quarter. On the first look it looks like the data does not follow any trend. Source: Author's illustration of Philadelphia Fed SPF data.

Horizon	HertzPersistence
0.25h	0.01
0.75h	0.03
1h	0.04
2h	0.06
5h	0.11
10h	0.15

Table 5: sMAPE of the author's persistence forecast with 50Hertz data for different forecast horizons. The error increases significantly with forecast horizon. Source: Author's illustration with data from 50Hertz and own persistence forecast, error calculated with Matlab

Horizon	50Hertz 2015	Tennet 2015
15/16h	0.05	0.06
20/21h	0.06	0.07
30/31h	0.07	0.08
38/39h	0.07	0.08

Table 6: sMAPE of 50Hertz and Tennet forecasts of 2015 and 2016 with different forecast horizons. The error increases slightly with forecast horizon. The increase is significantly smaller compared to the Philadelphia Fed SPF since the forecast are much more accurate for smaller horizons. Source: Author's illustration with data from netztransparenz.de (2017).

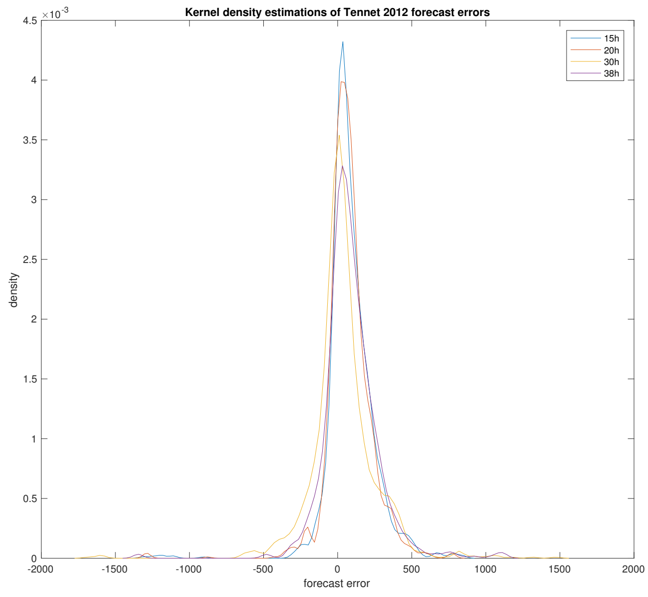

Figure 15: Kernel density estimation of Tennet 2012 wind power forecast error with various forecast horizons. With increasing forecast horizon, the conditional pdf gets less pronounced in its peak, the shoulders and tails do not change noticeable. Source: Author's illustration of German Wind Power data sets from netztransparenz.de (2017).

Variable	Mean	Standard deviation	Skewness	Kurtosis	Variable	Mean	Standard deviation	Skewness	Kurtosis
Hertz2012 15h	-51.96	174.25	0.24	9.58	USCPROF 1	0.48	37.61	-0.71	13.8
Hertz2012 20h	-65.18	190.37	1.47	32.78	USCPROF 2	0.95	58.16	-0.39	7.72
Hertz2012 30h	-55.14	214.06	-0.34	9.28	USCPROF 3	0.86	73.41	-0.26	5.78
Hertz2012 38h	-36.99	262.92	1.59	13.94	USCPROF 4	0.38	87.7	-0.47	4.94
Hertz2015 15h	-13.15	65.05	-4.04	71.21	USCPROF 5	0.56	99.63	-0.69	5.12
Hertz2015 20h	-9.85	57.74	-0.68	8.95	USHOUSING 1	0.02	0.1	-0.07	3.75
Hertz2015 30h	-7.67	72.43	0.19	5.21	USHOUSING 2	0.01	0.16	-0.14	3.66
Hertz2015 38h	-5.69	78.7	0.47	7.57	USHOUSING 3	-0.01	0.21	-0.4	3.53
Hertz2016 15h	-11.59	33.96	-0.04	6.38	USHOUSING 4	-0.03	0.25	-0.49	3.07
Hertz2016 20h	-11.11	40.84	1.21	14.99	USHOUSING 5	-0.05	0.29	-0.53	2.96
Hertz2016 30h	3.73	50.34	0.59	5.28	USINDPROD 1	-1.04	6.75	-6.8	55.42
Hertz2016 38h	-11.08	48.97	1.11	10.44	USINDPROD 2	-2.35	9.76	-4.34	24.55
P15	-0.01	13.78	0.39	19.71	USINDPROD 3	-3.81	12.12	-3.31	15.35
P30	-0.01	23.44	0.14	10.37	USINDPROD 4	-5.4	14.02	-2.69	10.86
P45	-0.02	31.66	0.07	8.06	USINDPROD 5	-7.05	15.62	-2.25	8.24
P1	-0.03	39.42	0.04	7.28	USNGDP 1	6.48	62.65	2.4	23.82
P2	-0.05	67.73	0.02	6.54	USNGDP 2	7.09	101.87	1.05	10.54
P5	-0.12	130.68	0.01	5.75	USNGDP 3	3.31	139.59	0.43	7.98
P10	-0.19	192.51	-0.06	5.05	USNGDP 4	-2.73	177.18	0.05	6.31
Tennet2012 16h	87.5	167.55	-1.51	21.27	USNGDP 5	-11.35	214.18	-0.33	5.51
Tennet2012 21h	77.04	154.29	-1.48	21.48	USUNEMP 1	-0.05	0.16	0.44	3.45
Tennet2012 31h	36.9	220.71	-0.23	15.43	USUNEMP 2	-0.06	0.38	1.09	5.06
Tennet2012 39h	89.88	194.04	0.12	15.11	USUNEMP 3	-0.03	0.57	1.45	6.53
Tennet2015 16h	63.04	98.07	1.61	7.92	USUNEMP 4	0.02	0.76	1.58	6.54
Tennet2015 21h	63.48	101.91	1.56	7.19	USUNEMP 5	0.08	0.94	1.47	5.7
Tennet2015 31h	79.3	127.72	1.52	5.86					
Tennet2015 39h	71.24	139.64	2.65	17.48					
Tennet2016 16h	39.11	61.69	1.73	7.76					
Tennet2016 21h	40.29	60.73	1.54	7.47					
Tennet2016 31h	52.79	80.13	1.46	5.57					
Tennet2016 39h	44.56	80.76	1.79	9.71					

Table 7: This table shows the statistical moments mean, standard deviation, skewness and kurtosis for all presented variables. Source: Author's illustration with data from netztransparenz.de 2017 and Philadelphia Fed SPF.

Variable	Test result	p-value	Variable	Test result	p-value	Variable	Test result	p-value
Hertz2012 15h	1	0	Tennet2012 16h	1	0	USHOUSING 3h	1	4.51E-19
Hertz2012 20h	1	0	Tennet2012 21h	1	0	USHOUSING 4h	1	2.34E-18
Hertz2012 30h	1	0	Tennet2012 31h	1	0	USHOUSING 5h	1	5.74E-17
Hertz2012 38h	1	0	Tennet2012 39h	1	0	USINDPROD 1h	0	0.1431
Hertz2015 15h	1	0	Tennet2015 16h	1	0	USINDPROD 2h	1	9.30E-12
Hertz2015 20h	1	0	Tennet2015 21h	1	0	USINDPROD 3h	1	2.52E-22
Hertz2015 30h	1	0	Tennet2015 31h	1	0	USINDPROD 4h	1	4.58E-33
Hertz2015 38h	1	0	Tennet2015 39h	1	0	USINDPROD 5h	1	2.51E-43
Hertz2016 15h	1	0	Tennet2016 16h	1	0	USNGDP 1h	1	4.20E-53
Hertz2016 20h	1	0	Tennet2016 21h	1	0	USNGDP 2h	1	6.26E-47
Hertz2016 30h	1	3.04E-290	Tennet2016 31h	1	0	USNGDP 3h	1	1.74E-46
Hertz2016 38h	1	0	Tennet2016 39h	1	0	USNGDP 4h	1	4.52E-47
Persistence 1h	1	0	USCPROF 1h	1	2.70E-26	USNGDP 5h	1	7.69E-55
Persistence 10h	1	0	USCPROF 2h	1	2.97E-33	USUNEMP 1h	1	2.88E-21
Persistence 15min	1	0	USCPROF 3h	1	4.12E-36	USUNEMP 2h	1	4.34E-12
Persistence 2h	1	0	USCPROF 4h	1	1.05E-41	USUNEMP 3h	1	9.62E-08
Persistence 30min	1	0	USCPROF 5h	1	1.04E-41	USUNEMP 4h	1	0.0000272
Persistence 45min	1	0	USHOUSING 1h	1	8.49E-29	USUNEMP 5h	1	0.00284

Table 8: Kolmogorov Smirnoff test results with p-values for all data sets. The null hypothesis of the data following a standard normal distribution is rejected for significance level alpha = 0.05 with very low p-values in all cases except for US industrial production index with a forecast horizon of 1h. These results confirm that our data series are typically not normal distributed. Source: Author's illustration with data from netztransparenz.de 2017 and Philadelphia Fed SPF.

Variable	Test result	p-value	Variable	Test result	p-value	Variable	Test result	p-value
Hertz2012 15h	1	0	Tennet2012 16h	1	0	USHOUSING 3h	1	0
Hertz2012 20h	1	0	Tennet2012 21h	1	0	USHOUSING 4h	1	0
Hertz2012 30h	1	0	Tennet2012 31h	1	0	USHOUSING 5h	1	0
Hertz2012 38h	1	0	Tennet2012 39h	1	0	USINDPROD 1h	0	0.9994
Hertz2015 15h	1	0	Tennet2015 16h	1	0	USINDPROD 2h	1	2.14E-6
Hertz2015 20h	1	0	Tennet2015 21h	1	0	USINDPROD 3h	1	0
Hertz2015 30h	1	0	Tennet2015 31h	1	0	USINDPROD 4h	1	0
Hertz2015 38h	1	0	Tennet2015 39h	1	0	USINDPROD 5h	1	0
Hertz2016 15h	1	0	Tennet2016 16h	1	0	USNGDP 1h	1	0.0142
Hertz2016 20h	1	0	Tennet2016 21h	1	0	USNGDP 2h	1	2.26E-14
Hertz2016 30h	1	0	Tennet2016 31h	1	0	USNGDP 3h	1	0
Hertz2016 38h	1	0	Tennet2016 39h	1	0	USNGDP 4h	1	0
Persistence 1h	1	0	USCPROF 1h	0	0.0586	USNGDP 5h	1	0
Persistence 10h	1	0	USCPROF 2h	1	0	USUNEMP 1h	0	0.3149
Persistence 15min	1	0	USCPROF 3h	1	0	USUNEMP 2h	1	1.12E-7
Persistence 2h	1	0	USCPROF 4h	1	0	USUNEMP 3h	1	0
Persistence 30min	1	0	USCPROF 5h	1	0	USUNEMP 4h	1	0
Persistence 45min	1	0	USHOUSING 1h	1	0.0071	USUNEMP 5h	1	0

Table 9: Ljung-Box Q-test results with p-values for all data sets. The null hypothesis that a time series exhibits no autocorrelation is rejected with very low p-values in almost every case. These results confirm that our data are typically significantly autocorrelated. Source: Author's illustration with data from netztransparenz.de 2017 and Philadelphia Fed SPF.

CPROF	1	2	3	4	5
1	1	0.74	0.57	0.52	0.46
2	0.74	1	0.85	0.73	0.66
3	0.57	0.85	1	0.89	0.8
4	0.52	0.73	0.89	1	0.92
5	0.46	0.66	0.8	0.92	1

HOUSING	1	2	3	4	5
1	1	0.56	0.38	0.3	0.36
2	0.56	1	0.73	0.59	0.52
3	0.38	0.73	1	0.83	0.72
4	0.3	0.59	0.83	1	0.87
5	0.36	0.52	0.72	0.87	1

INDPR	1	2	3	4	5
1	1	0.72	0.58	0.5	0.44
2	0.72	1	0.82	0.71	0.63
3	0.58	0.82	1	0.87	0.78
4	0.5	0.71	0.87	1	0.9
5	0.44	0.63	0.78	0.9	1

NGDP	1	2	3	4	5
1	1	0.74	0.6	0.51	0.46
2	0.74	1	0.86	0.75	0.68
3	0.6	0.86	1	0.91	0.83
4	0.51	0.75	0.91	1	0.94
5	0.46	0.68	0.83	0.94	1

UNEMP	1	2	3	4	5
1	1	0.65	0.51	0.39	0.32
2	0.65	1	0.83	0.69	0.56
3	0.51	0.83	1	0.9	0.78
4	0.39	0.69	0.9	1	0.93
5	0.32	0.56	0.78	0.93	1

Table 10: Pearsons crosscorrelation coefficient results. Source: Author's illustration with data from Philadelphia Fed SPF.

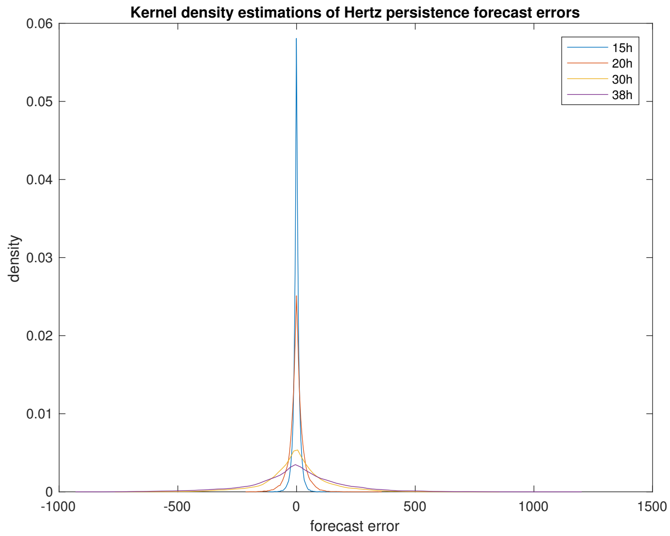

Figure 16: Kernel density estimation of 50Hertz 2015 wind power persistence forecast error with various forecast horizons. With increasing forecast horizon, the conditional pdf gets less pronounced in its peak, the shoulders get flatter and the tails get heavier. Source: Author's illustration of German Wind Power data sets from netztransparenz.de (2017).

	P15	P30	P45	P1	P2	P5	P10
P15	1	0.85	0.74	0.68	0.56	0.36	0.2
P30	0.85	1	0.92	0.84	0.7	0.45	0.26
P45	0.74	0.92	1	0.95	0.79	0.52	0.3
P1	0.68	0.84	0.95	1	0.86	0.58	0.34
P2	0.56	0.7	0.79	0.86	1	0.75	0.46
P5	0.36	0.45	0.52	0.58	0.75	1	0.74
P10	0.2	0.26	0.3	0.34	0.46	0.74	1

Table 11: Peasons crosscorrelation coefficient results. Source: Author's illustration with data from netztransparenz.de 2017.

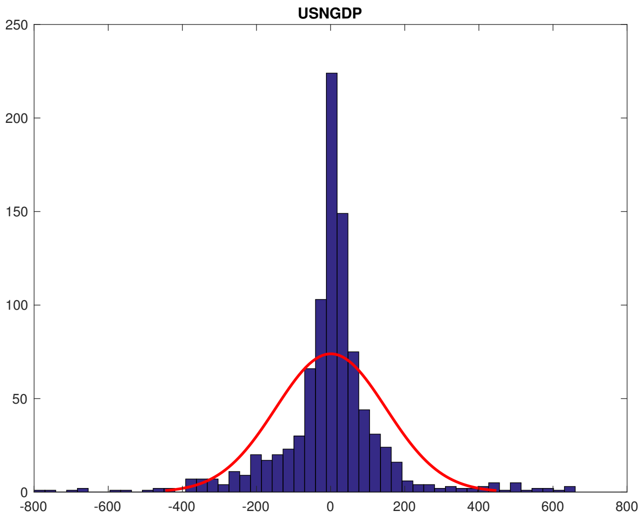

Figure 17: Histogram of US nominal GDP error with a Gaussian overlay of the same mean and variance. Note that this includes several combined forecast horizons, one to five quarters. The data has a mean of 0.56, a variance of 22164.29, a skewness of -0.09 and a kurtosis of 9.12. A visual examination of the conditional pdf shows more pronounced peaks, steeper shoulders and slightly fatter right tails than the Gaussian pdf. Source: Author's illustration of Philadelphia Fed SPF data.

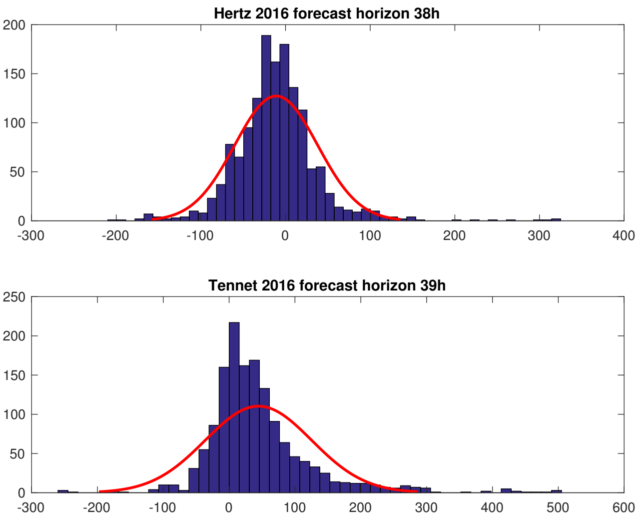

Figure 18: Histogram of 50Hertz and Tennet wind power forecast errors with a Gaussian overlay of the same mean and variance. The forecast horizon is 38 hours for 50Hertz and 39 hours for Tennet. The 50Hertz data has a mean of -11.08, a variance of 2398.44, a skewness of 1.11 and a kurtosis of 10.44. The Tennet data has a mean of 44.56, a variance of 6521.50, a skewness of 1.79 and a kurtosis of 9.71. A visual examination of the conditional pdf's shows more pronounced peak, steeper shoulders and heavier tails compared to the respective Gaussian distribution. Source: Author's illustration of netztransparenz.de 2017 data.

2012	Hertz 15h	Hertz 20h	Hertz 30h	Hertz 38h	Tennet 16h	Tennet 21h	Tennet 31h	Tennet 39h
Hertz 15h	1	0.55	0.12	0.08	0.26	0.17	0.11	-0.01
Hertz 20h	0.55	1	0.22	0.13	0.29	0.19	0.03	-0.01
Hertz 30h	0.12	0.22	1	0.18	0.01	-0.07	0.16	0.03
Hertz 38h	0.08	0.13	0.18	1	0	-0.07	0.19	0.34
Tennet 16h	0.26	0.29	0.01	0	1	0.7	0.36	0.29
Tennet 21h	0.17	0.19	-0.07	-0.07	0.7	1	0.39	0.29
Tennet 31h	0.11	0.03	0.16	0.19	0.36	0.39	1	0.39
Tennet 39h	-0.01	-0.01	0.03	0.34	0.29	0.29	0.39	1

2015	Hertz 15h	Hertz 20h	Hertz 30h	Hertz 38h	Tennet 16h	Tennet 21h	Tennet 31h	Tennet 39h
Hertz 15h	1	0.46	0.15	0.12	0.19	0.06	-0.03	0.04
Hertz 20h	0.46	1	0.34	0.17	0.11	0.19	0.09	0.01
Hertz 30h	0.15	0.34	1	0.28	-0.01	0.02	0.27	0.01
Hertz 38h	0.12	0.17	0.28	1	0.06	-0.01	0.09	0.45
Tennet 16h	0.19	0.11	-0.01	0.06	1	0.7	0.31	0.24
Tennet 21h	0.06	0.19	0.02	-0.01	0.7	1	0.46	0.27
Tennet 31h	-0.03	0.09	0.27	0.09	0.31	0.46	1	0.32
Tennet 39h	0.04	0.01	0.01	0.45	0.24	0.27	0.32	1

Table 12: Pearsons crosscorrelation coefficient results. Source: Author's illustration with data from netztransparenz.de 2017.

Variable	Test result	p-value	Variable	Test result	p-value	Variable	Test result	p-value
Hertz2012 15h	1	0	Tennet2012 16h	1	0	USHOUSING 3h	1	8.08E-13
Hertz2012 20h	1	0	Tennet2012 21h	1	0	USHOUSING 4h	1	0
Hertz2012 30h	1	0	Tennet2012 31h	1	0	USHOUSING 5h	1	0
Hertz2012 38h	1	0	Tennet2012 39h	1	0	USINDPROD 1h	0	0.8157
Hertz2015 15h	1	2.42E-11	Tennet2015 16h	1	0	USINDPROD 2h	1	4.59E-11
Hertz2015 20h	1	0	Tennet2015 21h	1	0	USINDPROD 3h	1	0
Hertz2015 30h	1	0	Tennet2015 31h	1	0	USINDPROD 4h	1	0
Hertz2015 38h	1	0	Tennet2015 39h	1	0	USINDPROD 5h	1	0
Hertz2016 15h	1	0	Tennet2016 16h	1	0	USNGDP 1h	0	0.957
Hertz2016 20h	1	0	Tennet2016 21h	1	0	USNGDP 2h	1	4.29E-11
Hertz2016 30h	1	0	Tennet2016 31h	1	0	USNGDP 3h	1	2.33E-15
Hertz2016 38h	1	0	Tennet2016 39h	1	0	USNGDP 4h	1	0
Persistence 1h	1	0	USCPROF 1h	0	0.7788	USNGDP 5h	1	0
Persistence 10h	1	0	USCPROF 2h	1	4.98E-10	USUNEMP 1h	0	0.1001
Persistence 15min	1	0	USCPROF 3h	1	2.36-14	USUNEMP 2h	1	0.00000145
Persistence 2h	1	0	USCPROF 4h	1	0	USUNEMP 3h	1	1.52E-13
Persistence 30min	1	0	USCPROF 5h	1	0	USUNEMP 4h	1	0
Persistence 45min	1	0	USHOUSING 1h	1	0.01898	USUNEMP 5h	1	0

Table 13: ARCH test results with p-values for all data sets. Source: Author's illustration with data from netztransparenz.de 2017 and Philadelphia Fed SPF.

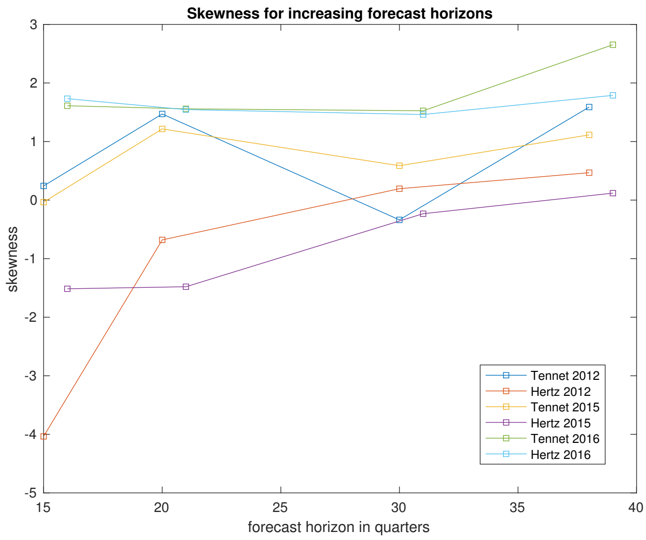

Figure 19: Skewness of various Wind power forecasts in Germany by TSO with increasing forecast horizons. The skewness here tends to have a slight increasing trend. Source: Author's illustration of netztransparenz.de 2017 data.

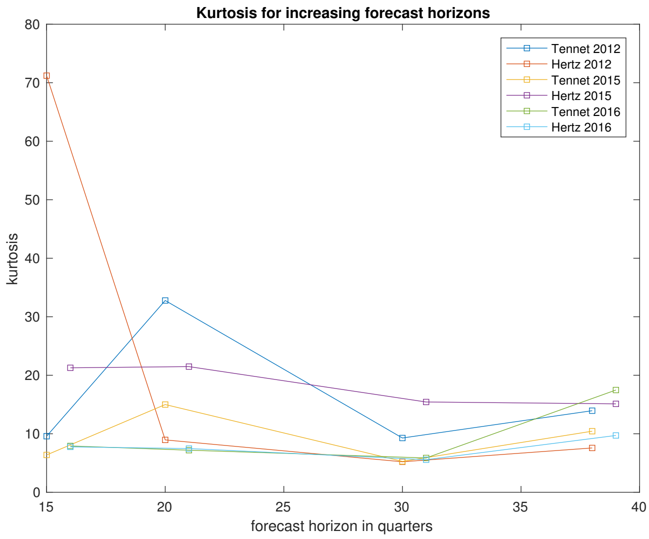

Figure 20: Kurtosis of various Wind power forecasts in Germany by TSO with increasing forecast horizons. The values reach from 5.21 to 71.21. The kurtosis here has a slightly decreasing trend with increasing forecast horizon. The decrease is not so clear as the forecast horizons are not in a wide range of values. Source: Author's illustration of netztransparenz.de 2017 data.

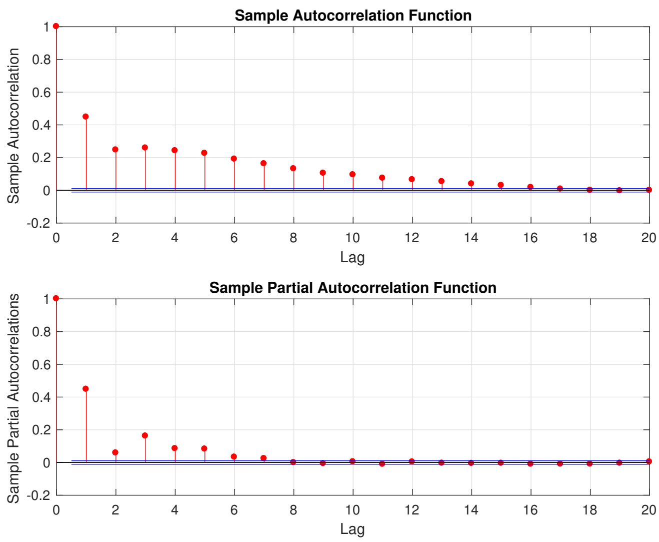

Figure 21: Autocorrelation P15. Source: Author's illustration of netztransparenz.de 2017 data.

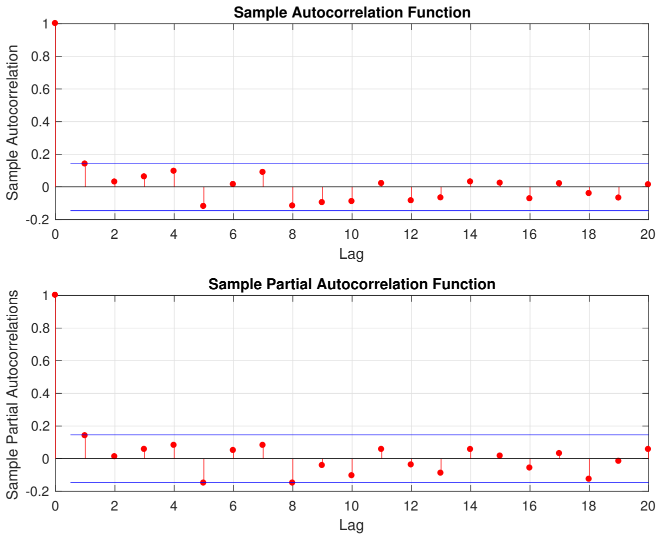

Figure 22: Autocorrelation USUNEMP1. Source: Author's illustration of Philadelphia Fed SPF data.

B Statistical tests

In this work various statistical hypothesis tests are used. They are defined below.

B.1 Augmented Dickey-Fuller test

The *Augmented Dickey-Fuller test* is a popular unit root test. The test assesses the null hypothesis of a unit root and uses the expanded autoregressive model in equation 27.

$$y_t = c + \delta t + \varphi y_{t-1} + \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + \epsilon_t \tag{27}$$

y_t is the time series, $c + \delta t$ is a time trend, $\beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p}$ are lagged values, p is the number of lagged difference terms, which default value is 0 and ϵ_t is mean zero error process. Equation 28 shows the null and typical alternative hypothesis that the time series is stationary.

$$H_0 : \varphi = 1, \quad H_1 : \varphi < 1 \tag{28}$$

Equation 29 shows the test stastic, which calculates a OLS estimation of $\hat{\varphi}$ and its standard error se.

$$t_{adf} = \frac{(\hat{\varphi} - 1)}{se} \tag{29}$$

Critical values are calculated using Monte-Carlo simulation. It is important to note that, for small sample sizes, the critical values are only valid if the input is Gaussian distributed. Results for non-Gaussian input is valid for large sample sizes. In this work the standard significance level is applied, so *alpha* is 0.05 (Mathworks Documentation 2017c; Ruppert 2011h).

B.2 Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test

The *KPSS test* assesses the null hypothesis of trend stationarity and uses the model in equation 30.

$$y_t = c_t + \delta t + u_{1t}, \quad c_t = c_{t-1} + u_{2t} \tag{30}$$

y_t is the time series, c_t is a random walk term, δ is the trend coefficient, u_{1t} is a stationary process and u_{2t} is an independent and identically distributed (iid) process. Equation 31 shows the null and alternative hypothesis of a nonstationary unit root process. The null hypothesis means that c_t is constant and the alternative hypothesis a unit root process in the random walk.

$$H_0 : \sigma^2_{u_{2t}} = 0, \quad H_1 : \sigma^2_{u_{2t}} > 0 \tag{31}$$

Equation 29 shows the test stastic, where T is the sample size, s^2 is the Newey-West estimate of the long-run variance and $\sum_{t=1}^{T} S_t^2$ is the sum of squared residuals of the regression of y on an intercept and time trend.

$$t_{kpss} = \frac{\sum_{t=1}^{T} S_t^2}{s^2 T^2} \tag{32}$$

Critical values are calculated using Monte-Carlo simulation. In this work the standard significance level is applied, so *alpha* is 0.05 (Kwiatkowski et al. 1992; Mathworks Documentation 2017f).

B.3 Kolmogorov-Smirnov test

The *Kolmogorov-Smirnov test* assesses the null hypothesis that a time series follows a hypothesized cumulative distribution function (cdf), in this work a standard normal distribution.

$$t_{ks} = \max_x(|\hat{F}(x) - G(x)|) \tag{33}$$

Equation 33 shows the test stastic, where $\hat{F}(x)$ is the empirical cdf and $G(x)$ is the cdf of the standard normal distribution. In this work the standard significance level is applied, so *alpha* is 0.05 (Mathworks Documentation 2017g).

B.4 Ljung-Box Q-test

The *Ljung-Box Q test* assesses the null hypothesis that a time series's autocorrelation is zero in the first m lags. Equation 34 shows the null and alternative hypothesis of a time series exhibiting autocorrelation.

$$H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0, \quad H_1 : \rho_1 = \rho_2 = ... = \rho_m \neq 0 \tag{34}$$

Equation 35 shows the test stastic, where N is the length of the time series, m the number of tested lags and $\hat{\rho}_h$ the sample autocorrelation value for lag h.

$$t_{lbq} = N(N+2) \sum_{h=1}^{m} \frac{\hat{\rho}_h^2}{N-h} \tag{35}$$

The test statistic t_{lbq} follows a χ^2_m distribution under the null hypothesis, which is rejected if $t_{lbq} > \chi^2_{1-\alpha,m}$ The time series tested must have a constant mean, thus be detrended. In this work the standard significance level is applied, so *alpha* is 0.05 (Mathworks Documentation 2017h; Ljung and Box 1978).

B.5 Engle's ARCH test

Engle's ARCH test assesses the null hypothesis that a time series exhibits no conditional heteroscedasticity (*ARCH effects*) using the *ARCH(L)* model in equation 36. r_t is a time series, e_t are zero mean error and L is the number of lagged terms.

$$r_t^2 = a_0 + a_1 r_{t-1}^2 + ... + a_L r_{t-L}^2 + e_t \tag{36}$$

Equation 37 shows the null and alternative hypothesis, that an *ARCH(L)* in equation 36 describes the time series, that means there is at least one significant coefficient a.

$$H_0 : a_1 = a_2 = ... = a_L = 0, \quad H_1 : \exists a_j \neq 0; \quad j = 1, 2, ..., L \tag{37}$$

The test fits the *ARCH(L)* model in equation 36 on the time series, then regresses the squared error terms e_t on a constant and L lags and then calculates the test statistic in equation 38, where T is the sample size and R^2 is the coefficient of determination from the regression of the squared error terms.

$$t_{arch} = TR^2 \tag{38}$$

The test statistic t_{arch} follows a χ^2_L distribution under the null hypothesis, which is rejected if $t_{arch} > \chi^2_{1-\alpha,L}$. In this work the standard significance level is applied, so *alpha* is 0.05 (Mathworks Documentation 2017d; Engle 1982).

B.6 Johansen cointegration test

The *Johansen cointegration test* is a test for *cointegration* relationships in a multivariate time series and estimates the VECM coefficients via maximum likelihood estimation. There are two types of tests, one with *trace* and one with *eigenvalues*. We present the *trace* test in this work, as it is the default value of the implemented version in *Matlab*. There is no fundamental difference in the result, only in the formulation of the hypotheses.

Equation 39 shows a VECM, where $\Pi = \alpha\beta^T$, μ is a mean vector and D_t a vector of seasonal dummies.

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + ... + \Gamma_{p-1} \Delta Y_{t-p+1} + \mu + \Phi D_t + \epsilon_t \tag{39}$$

Let us recall from section 3.4 that α is the loading matrix and β the cointegration matrix. The columns of β represent the cointegration relationships. The test assesses the rank of Π, which is the number of cointegration vectors, so the number of independent linear combinations. It is a sequential procedure, it iterates through $r_0 = 0, 1, ..., K-1$, and stops at the first r_0 for which the null hypothesis can not be rejected. This rank r_0 is the number of cointegration relationships. The hypotheses are shown in equation 40, where the null hypothesis means that the cointegration rank is the current tested rank, while a rejection means that there are more cointegration relationships than currently tested. K is the number of time series.

$$H_0(r_0) : rank(\Pi) = r_0, \quad H_1(r_0) : rank(\Pi) > r_0; \quad r_0 = 0, 1, ..., K-1 \quad (40)$$

There are two special cases, the first is $r_0 = 0$, which means that there no cointegrating relationships. The second is $r_0 = K - 1$, which means that the variables are already stationary, thus a VECM makes no sense. Equation 41 shows the test statistic, where $l(r_0)$ is the maximum of the Gaussian likelihood function and λ_i is the i^{th} eigenvalue obtained by the coefficient estimation.

$$t_{jci,trace}(r_0) = 2(\log l(r_0) - \log l(K)) = -T \sum_{i=r_0+1}^{K} \log(1 - \lambda_i) \quad (41)$$

For the maximum likelihood estimation of coefficients and estimation of critical values we refer to Mathworks Documentation (2017e), Johansen (1991), and Kilian and Lütkepohl (2016), where also the definitions and notations of this section come from.

References

50Hertz Transmission GmbH (2017). *Forecast wind power feed-in.* URL: `http://www.50hertz.com/en/Grid-Data/Wind-power/Forecast-wind-power-feed-in` (visited on 02/22/2017).

Armstrong, J Scott and Fred Collopy (1992). "Error measures for generalizing about forecasting methods: Empirical comparisons". In: *International journal of forecasting* 8.1, pp. 69–80.

Auer, Benjamin and Horst Rottmann (2011). *Statistik und okonometrie fur wirtschaftswissenschaftler: eine...* Springer, pp. 613–616.

Bahmani-Oskooee, Mohsen and Janardhanan Alse (1994). "Short-run versus long-run effects of devaluation: error-correction modeling and cointegration". In: *Eastern Economic Journal* 20.4, pp. 453–464.

Balanda, Kevin P and HL MacGillivray (1988). "Kurtosis: a critical review". In: *The American Statistician* 42.2, pp. 111–119.

Bludszuweit, Hans, José Antonio Domínguez-Navarro, and Andrés Llombart (2008). "Statistical analysis of wind power forecast error". In: *IEEE Transactions on Power Systems* 23.3, pp. 983–991.

Bofinger, S, A Luig, and HG Beyer (2002). "Qualification of wind power forecasts". In: *2002 Global Windpower Conference, Paris.* Vol. 2. 5.

Brockwell, Peter J and Richard A Davis (2016a). *Introduction to time series and forecasting.* springer.

— (2016b). *Introduction to time series and forecasting.* springer.

Coibion, Olivier and Yuriy Gorodnichenko (2008). *What can survey forecasts tell us about informational rigidities?* Tech. rep. National Bureau of Economic Research.

Cont, Rama (2001). "Empirical properties of asset returns: stylized facts and statistical issues". In: *Quantitative Finance* 1.2, pp. 223–236.

— (2005). "Long range dependence in financial markets". In: *Fractals in Engineering.* Springer, pp. 159–179.

Cont, Rama (2007). "Volatility clustering in financial markets: empirical facts and agent-based models". In: *Long memory in economics*. Springer, pp. 289–309.

Croushore, Dean D (1993). "Introducing: the survey of professional forecasters". In: *Business Review-Federal Reserve Bank of Philadelphia* 6, p. 3.

DeCarlo, Lawrence T (1997). "On the meaning and use of kurtosis." In: *Psychological methods* 2.3, p. 292.

Diebold, Francis X and Jose A Lopez (1996). "8 Forecast evaluation and combination". In: *Handbook of statistics* 14, pp. 241–268.

Dovern, Jonas and Johannes Weisser (2011). "Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7". In: *International Journal of Forecasting* 27.2, pp. 452–465.

Elliott, Graham and Allan Timmermann (2004). "Optimal forecast combinations under general loss functions and forecast error distributions". In: *Journal of Econometrics* 122.1, pp. 47–79.

Engle, Robert F (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Engle, Robert F and Clive WJ Granger (1987). "Co-integration and error correction: representation, estimation, and testing". In: *Econometrica: journal of the Econometric Society*, pp. 251–276.

Fabbri, Alberto et al. (2005). "Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market". In: *IEEE Transactions on Power Systems* 20.3, pp. 1440–1446.

Federal Reserve Bank of Philadelphia (2017a). *Historical Data Files for the Survey of Professional Forecasters*. URL: `https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/ngdp` (visited on 04/12/2017).

Federal Reserve Bank of Philadelphia (2017b). *Real-Time Data Set: Full-Time Series History.* URL: https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files (visited on 04/12/2017).

— (2017c). *Survey of Professional Forecasters.* URL: https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/ (visited on 02/22/2017).

— (2017d). *Survey of professional forecasters documentation.* URL: https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf?la=en (visited on 04/16/2017).

Florita, A., B.M. Hodge, and M.R. Milligan (2012). *Wind Power Forecasting Error Frequency Analyses for Operational Power System Studies: Preprint.* Conference paper. URL: https://books.google.de/books?id=ot74vQAACAAJ.

Focken, Ulrich et al. (2002). "Short-term prediction of the aggregated power output of wind farms,Äîa statistical analysis of the reduction of the prediction error by spatial smoothing effects". In: *Journal of Wind Engineering and Industrial Aerodynamics* 90.3, pp. 231–246.

Granger, Clive WJ (1981). "Some properties of time series data and their use in econometric model specification". In: *Journal of econometrics* 16.1, pp. 121–130.

Granger, Clive WJ and Paul Newbold (1974). "Spurious regressions in econometrics". In: *Journal of econometrics* 2.2, pp. 111–120.

Granger, Clive WJ and M Hashem Pesaran (2004). "Economic and statistical measures of forecast accuracy". In:

Hodge, Bri-Mathias, Erik Ela, and Michael Milligan (2011). "The distribution of wind power forecast errors from operational systems". In: *10th International Workshop on Large-scale Integration of Wind Power.*

Hodge, Bri-Mathias, Anthony Florita, et al. (2012). "A comparison of wind power and load forecasting error distributions". In: *2012 World Renewable Energy Forum*, pp. 13–17.

Hodge, Bri-Mathias, Debra Lew, et al. (2012). "Wind power forecasting error distributions: An international comparison". In: *11th Annual International Workshop on Large-Scale Integration of Wind Power into Power Systems as well as on Transmission Networks for Offshore Wind Power Plants Conference*.

Hodge, Bri-Mathias and Michael Milligan (2011). "Wind power forecasting error distributions over multiple timescales". In: *2011 IEEE power and energy society general meeting*. IEEE, pp. 1–8.

Holttinen, H, TS Nielsen, and G Giebel (2002). "Wind energy in the liberalised market–forecast errors in a day-ahead market compared to a more flexible market mechanism". In: *Proceeding, International Symposium on Distributed Generation: Power System and Market Aspects, Stockholm*.

Hyndman, Rob J and Anne B Koehler (2006). "Another look at measures of forecast accuracy". In: *International journal of forecasting* 22.4, pp. 679–688.

Johansen, Søren (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". In: *Econometrica: Journal of the Econometric Society*, pp. 1551–1580.

Kaldor, Nicholas (1961). "Capital accumulation and economic growth". In: *The theory of capital*. Springer, pp. 177–222.

Kilian, Lutz and Helmut Lütkepohl (2016). *Structural Vector Autoregressive Analysis*.

Kwiatkowski, Denis et al. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" In: *Journal of econometrics* 54.1-3, pp. 159–178.

Lange, Matthias (2003). “Analysis of the uncertainty of wind power predictions”. PhD thesis. Universität Oldenburg.

— (2005). “On the uncertainty of wind power predictions-Analysis of the forecast accuracy and statistical distribution of errors”. In: *Journal of Solar Energy Engineering* 127.2, pp. 177–184.

Lee, Dongin and Peter Schmidt (1996). “On the power of the KPSS test of stationarity against fractionally-integrated alternatives”. In: *Journal of econometrics* 73.1, pp. 285–302.

Ljung, Greta M and George EP Box (1978). “On a measure of lack of fit in time series models”. In: *Biometrika*, pp. 297–303.

Luig, Armin, Stefan Bofinger, and Hans Georg Beyer (2001). “Analysis of confidence intervals for the prediction of regional wind power output”. In: *Proceedings of the European wind energy conference, Copenhagen, Denmark*, pp. 725–728.

Madsen, Henrik et al. (2005). “Standardizing the performance evaluation of short-term wind power prediction models”. In: *Wind Engineering* 29.6, pp. 475–489.

Makridakis, Spyros, A Andersen, et al. (1982). “The accuracy of extrapolation (time series) methods: Results of a forecasting competition”. In: *Journal of forecasting* 1.2, pp. 111–153.

Makridakis, Spyros, Chris Chatfield, et al. (1993). “The M2-competition: A real-time judgmentally based forecasting study”. In: *International Journal of Forecasting* 9.1, pp. 5–22.

Makridakis, Spyros and Michele Hibon (2000). “The M3-Competition: results, conclusions and implications”. In: *International journal of forecasting* 16.4, pp. 451–476.

Makridakis, Spyros, Michèle Hibon, et al. (1987). “Confidence intervals: An empirical investigation of the series in the M-competition”. In: *International Journal of Forecasting* 3.3-4, pp. 489–508.

Masih, Abul MM and Rumi Masih (1997). "On the temporal causal relationship between energy consumption, real income, and prices: some new evidence from Asian-energy dependent NICs based on a multivariate cointegration/vector error-correction approach". In: *Journal of policy modeling* 19.4, pp. 417–440.

Mathworks Documentation (2017a). *Mathworks Documentation*. URL: `https://de.mathworks.com/help/econ/index.html` (visited on 04/26/2017).

— (2017b). *Mathworks Documentation, vec2var*. URL: `https://de.mathworks.com/help/econ/vec2var.html` (visited on 04/26/2017).

— (2017c). *Matlab adftest*. URL: `https://de.mathworks.com/help/econ/adftest.html` (visited on 04/27/2017).

— (2017d). *Matlab archtest*. URL: `https://de.mathworks.com/help/econ/archtest.html` (visited on 04/27/2017).

— (2017e). *Matlab jcitest*. URL: `https://de.mathworks.com/help/econ/jcitest.html` (visited on 04/26/2017).

— (2017f). *Matlab kpsstest*. URL: `https://de.mathworks.com/help/econ/kpsstest.html` (visited on 04/27/2017).

— (2017g). *Matlab kstest*. URL: `https://de.mathworks.com/help/stats/kstest.html` (visited on 04/27/2017).

— (2017h). *Matlab lbqtest*. URL: `https://de.mathworks.com/help/econ/lbqtest.html` (visited on 04/27/2017).

Mello, Phillip E, Ning Lu, and Yuri Makarov (2011). "An optimized autoregressive forecast error generator for wind and load uncertainty study". In: *Wind Energy* 14.8, pp. 967–976.

Meyler, Aidan and Ieva Rubene (2009). "Results of a special questionnaire for participants in the ECB Survey of Professional Forecasters (SPF)". In:

Mukherjee, Tarun K and Atsuyuki Naka (1995). "Dynamic relations between macroeconomic variables and the Japanese stock market: an application of a vector error correction model". In: *Journal of Financial Research* 18.2, pp. 223–237.

netztransparenz.de (2017). *netztransparenz.de.* URL: `https://www.netztransparenz.de` (visited on 02/22/2017).

Pinson, Pierre (2006). "Estimation of the uncertainty in wind power forecasting". PhD thesis. École Nationale Supérieure des Mines de Paris.

Pinson, Pierre and Georges Kariniotakis (2004). "On-line assessment of prediction risk for wind power production forecasts". In: *Wind Energy* 7.2, pp. 119–132.

Rossi, Barbara and Tatevik Sekhposyan (2015). "Macroeconomic uncertainty indices based on nowcast and forecast error distributions". In: *The American Economic Review* 105.5, pp. 650–655.

Ruppert, David (2011a). *Statistics and data analysis for financial engineering.* Springer, pp. 201–208.

— (2011b). *Statistics and data analysis for financial engineering.* Springer, pp. 354–361.

— (2011c). *Statistics and data analysis for financial engineering.* Springer, pp. 413–422.

— (2011d). *Statistics and data analysis for financial engineering.* Springer, pp. 235, 270–276.

— (2011e). *Statistics and data analysis for financial engineering.* Springer, pp. 81–86.

— (2011f). *Statistics and data analysis for financial engineering.* Springer, pp. 44–48.

— (2011g). *Statistics and data analysis for financial engineering.* Springer, pp. 477–502.

— (2011h). *Statistics and data analysis for financial engineering.* Springer, pp. 233–236.

Stark, Tom et al. (2010). "Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters". In: *Federal Reserve Bank of Philadelphia Research Rap, Special Report* 1.

Tewari, Saurabh, Charles J Geyer, and Ned Mohan (2011). "A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets". In: *IEEE Transactions on Power Systems* 26.4, pp. 2031–2039.

Xcel Energy Inc. (2017a). *Wind Operations*. URL: `https://www.xcelenergy.com/energy_portfolio/renewable_energy/wind/wind_operations` (visited on 01/10/2017).

— (2017b). *Xcel Energy - About us*. URL: `https://www.xcelenergy.com/company/about_us` (visited on 04/26/2017).