House of Energy Markets & Finance

# FORECASTING THE DISTRIBUTIONS OF HOURLY ELECTRICITY SPOT PRICES

## – ACCOUNTING FOR SERIAL CORRELATION PATTERNS AND NON-NORMALITY OF PRICE DISTRIBUTIONS

**HEMF Working Paper No. [05/2017]**

*by*

*Christian Pape,*

*Arne Vogler,*

*Oliver Woll*

*and*

*Christoph Weber*

*2017*

UNIVERSITÄT DUISBURG ESSEN

*Open-Minded*

EUROPÄISCHE UNION
Investition in unsere Zukunft
Europäischer Fonds für regionale Entwicklung

EFRE.NRW
Investitionen in Wachstum und Beschäftigung

Ministerium für Wirtschaft, Energie, Industrie, Mittelstand und Handwerk des Landes Nordrhein-Westfalen

# Forecasting the distributions of hourly electricity spot prices*

By Christian Pape**, Arne Vogler, Oliver Woll and Christoph Weber


## Abstract

We present a stochastic modelling approach to describe the dynamics of hourly electricity prices. The suggested methodology is a stepwise combination of several mathematical operations to adequately characterize the distribution of electricity spot prices. The basic idea is to analyze day-ahead prices as panel of 24 cross-sectional hours and to identify principal components of hourly prices to account for the cross correlation between hours. Moreover, non-normality of residuals is addressed by performing a normal quantile transformation and specifying appropriate stochastic processes for time series before fit. We highlight the importance of adequate distributional forecasts and present a framework to evaluate the distribution forecast accuracy. The application for German electricity prices 2015 reveal that: (i) An autoregressive specification of the stochastic component delivers the best distribution but not always the best point forecasting results. (ii) Only a complete evaluation of point, interval and density forecast, including formal statistical tests, can ensure a correct model choice.


*Keywords: Distribution forecasts, Electricity, Price forecasting, Panel data, Statistical tests*

*JEL-Classification: Q47, N74*

CHRISTIAN PAPE (**CORRESPONDING AUTHOR)
Ph. D. Student, House of Energy Markets and Finance
and Commercial Analyst, innogy SE
Gildehofst. 1, 45127 Essen
+49-(0)201 / 1214 - 373
christian.pape@uni-due.de
or christian.pape@innogy.com

ARNE VOGLER, arne.vogler@uni-due.de
CHRISTOPH WEBER, christoph.Weber@uni-due.de
University of Duisburg-Essen, Germany

OLIVER WOLL, woll@zew.de
Centre of European Economic Research (ZEW), Germany



**The research presented in this paper has been partially carried out within the research project 'StoOpt.NRW' financed by the Ministry of Economics NRW (MWEIMH NRW) and the European Regional Development Fund (EFRE), allocated by the European Union.*

## 1 Introduction

Short-term electricity price forecasting has gained importance as market participants in the energy industry are exposed to the risk of electricity price variations or try to profit from volatile prices. In particular, recently increasing shares of variable renewable energy production substantially influence the uncertainties associated with electricity price forecasting. In order to improve the quality of decision making, e. g. for risk management purposes, the use of stochastic bidding and dispatch models, reliable distribution forecasts of prices are important. Also for longer-term decision making, e. g. for a reasonable valuation of flexible assets or stochastic optimization models, it is essential to forecast every hour of a day consistently and to ensure realistic smooth transitions between consecutive hourly price predictions. The challenges of predicting hourly electricity prices are a result of the idiosyncratic influences on the price formation. Mainly caused by the almost non-storability of electricity as well as fundamental production and demand characteristics, the electricity spot price time series are exhibiting characteristics like: Daily, weekly and seasonal cycles, high and non-constant volatility (heteroscedasticity), mean reversion and frequent outliers (positive and negative spikes). Short-term electricity price forecasting has to represent these characteristics. In practice, a well-performing simulation procedure that keeps time and effort small is required.

Our approach belongs to the group of econometric-stochastic approaches and can be combined with different point forecasting methods. To highlight the benefits of the stochastic approach in this paper, we combine it with a relatively simple regression approach. The suggested method is a stepwise combination of several mathematical operations to adequately characterize the distributions of hourly electricity prices. One major element is a Principal Component Analysis (PCA) of hourly electricity prices by using an eigenvalue-eigenvector decomposition. Moreover, non-normality of residuals is accounted for by performing a normal quantile transformation. We consider several different specifications to capture autoregressive effects and time-varying volatility of the stochastic components.

We contribute to the existing literature by developing a stochastic approach for describing highly volatile time series that delivers reliable distribution forecasts. To the best of our knowledge, the suggested quantile mapping and identification of principal components for the cross correlation is new to the price forecasting literature.[1] Additionally, we highlight the importance of distribution forecasting. Distribution forecasting has featured intensively in the economic-financial literature and probabilistic

[1] Closely related works are from Härdle and Trueck (2010); Kovacevic and Wozabal (2014); Keles et al. (2013); Huisman and Kilic (2013).

electricity price forecasting has started to gain importance as of late (see Nowotarski and Weron 2016). Yet, said strand of electricity price forecasting requires further research and, to the best of our knowledge, no study has applied a distribution forecast evaluation framework as comprehensive as the one presented here. We present a framework to test the quality of distributional forecasts for electricity prices going beyond the evaluation of point and interval forecasts. The remainder of the paper is structured as follows. Chapter 2 describes the theoretical background for the stochastic modelling approach and the procedure to evaluate the forecast quality. In chapter 0, the presented methodology is applied to the German electricity spot prices. After presenting the estimation and forecast results for point, interval and distribution forecasts, we conclude on the forecast ability of different model specifications. Chapter 0 summarizes the article and provides a short outlook.

## 2 Methodology

### 2.1 Literature

Forecasting models for electricity prices are commonly classified by the planning horizon and the applied methodology. The strength of econometric-stochastic (top down) approaches is the modelling of price volatilities. Yet, due to the almost non-storability of electricity, the no-arbitrage assumption for derivative (futures) markets is only true in expectation. Consequently, econometric-stochastic approaches are not exclusively adequate. Fundamental (bottom-up) approaches are beneficial to consider longer-term changes in supply and demand, whereas these models struggle to reproduce price variations and volatilities. So called hybrid modelling approaches try to combine the benefits from different modelling methodologies (see Pape et al. 2017).

In recent years the forecasting of day-ahead electricity prices has been studied intensively. Weron (2014) delivers an extensive review of state-of-the-art electricity price forecasting methods. Other general works on electricity price modelling are done by Ventosa et al. (2005) summarizing electricity price modelling trends or a reviewing paper on this topic by Aggarwal et al. (2009). The literature on probabilistic electricity price forecasting has gained importance lately and continues to expand (Nowotarski and Weron 2016; Weron 2014). Given that well performing point forecasting models are not necessarily the best for distribution forecasting, a result underscored by the present study, the focus on probabilistic forecasting seems warranted. Thus, the correct choice of electricity price forecasting models requires a complete evaluation of point, interval and distribution forecasting ability.

Within the wide field of electricity price forecasting, some related works to this paper are the following: Janczura and Weron (2010) run and test goodness-of-fit for various Markov regime-switching models. They find the best structure for '[…] independent spike 3-regime model with time-varying transition probabilities, heteroscedastic diffusion-type base regime dynamics and shifted spike regime distributions.' Their application is on mean daily day-ahead prices from 2001 to 2005 for EEX, PJM Interconnection and the New England Power Pool. One work on hourly electricity spot prices is by Conejo et al. (2005) considering time series analysis, neural networks and wavelets to forecast electricity prices for the PJM market 2002. Checking the short and long term forecast ability, they conclude that time series are most efficacious and within time series those with dynamic regression and transfer function are more effective than ARIMA. Also forecasting within-day electricity prices, Karakatsani and Bunn (2008) consider an application to the UK half-hourly market. They highlight the importance to include fundamental factors within time series models for electricity price forecasting. Another related work on hourly spot electricity prices is by Weron and Misiorek (2008) comparing the accuracy of 12 time series models for day-ahead spot prices in Spain and California. They show that load as exogenous variable is essential for the quality of the models and that semi-parametric models generally lead to better point and interval forecasts under different market conditions. Jónsson et al. (2014) investigate density forecasts of energy prices for the Nord Pool market. They use a semi-parametric methodology to forecast densities for prices. Their model outperforms four benchmarks including a GARCH model by highlighting the ability to deliver reliable quantile estimates.

Works on hourly electricity price forecasting applied to the German market are the following: Kosater and Mosler (2006) compare different time series approaches with nonlinear Markov regime-switching to ordinary linear autoregressive models in terms of their forecast ability. They infer that nonlinear regime switching models deliver better results especially for longer-run models for the German spot market. Bierbrauer et al. (2007) compare various regime switching models to jump diffusion and mean reversion models. In their application regime switching approaches outperform their benchmarks. Härdle and Trueck (2010) suggest a dynamic semiparametric factor model (DSFM) for hourly electricity prices. They identify three factors explaining 80 percent of the variation in hourly Germany electricity prices from 2005 to 2008. Kovacevic and Wozabal (2014) forecast day-ahead electricity prices with '[…] a semi parametric, single-index, generalized linear model. PCA is used to reduce complexity of the model. Their results for EEX and PJM data show good forecast quality compared to simple benchmarks and a seasonal ARIMA model. The only exception from these good results are very low prices in the German market. In light of this, Keles et al. (2012) compare point forecasting results and the coverage of negative prices from different stochastic processes (mean reversion, ARMA, ARIMA and GARCH) that try to incorporate negative prices within a regime switching model. They highlight the necessity of regime switching models to cover price jumps (positive and negative) and improve the error measures. In their application to German electricity prices, the ARMA model delivers the lowest errors between actual and simulated prices. Recent works on hourly price forecasting highlight the increasing importance of renewable energy uncertainty for electricity price modelling in Germany (among others Keles et al. 2013; Pape et al. 2016; Wozabal et al. 2014; Ziel et al. 2015). With increasing shares of renewable production in today's power systems, electricity prices behave differently from those of fossil-fuel-based systems.

## 2.2 A stochastic forecasting approach

The approach in this article belongs to the class of econometric-stochastic approaches and focusses on explaining causal dependencies between factors influencing electricity price formation. Here, electricity prices are analyzed separately for every single hour in a so-called panel of 24 cross-sectional hours (Huisman et al. 2007).[2] The intuition behind the panel analysis is that day-ahead bidding is done simultaneously for all 24 hours of the same day and based on the same information set. Therefore, the 24 series are expected to be interrelated and dependent. In order to characterize distributions adequately, we suggest a combination of several mathematic approaches. The estimation and simulation procedure includes the following steps, whereby step (i), (ii), (v) and (vi) are adjustable depending on the application:

- (i) Treat the time series of spot prices as panel of different individual hours and transform the original values
- (ii) Determine the main deterministic influences and the residuals
- (iii) Map the residuals' empirical distribution to a normal distribution
- (iv) Identify the common factors of hourly prices
- (v) Model lagged effects of price level and price volatility
- (vi) Use rolling window technique to estimate and simulate

(i) The first step is the setup of vectors $\boldsymbol{p}_h$ for individual hours $h \in \{1, \dots, H\}$ observed with daily frequency $t \in \{1, \dots, T\}$ where each vector contains $T$ elements, i. e. the number of days in the dataset.

[2] E. g. Cuaresma et al. (2004) have shown that modelling each hour of the day separately outperforms models that calculate daily price time series.

$$\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_{24}) = \begin{pmatrix} p_{1,1} & p_{2,1} \cdots & p_{H,1} \\ p_{1,2} & p_{2,2} \cdots & p_{H,2} \\ \vdots & \vdots \ddots & \vdots \\ p_{1,T} & p_{2,T} \cdots & p_{H,T} \end{pmatrix} \tag{2-1}$$

A transformation of the original values is needed if the original time series is far away from normality. Applying a non-linear transformation like the logarithmic transformation to obtain normality is a common procedure in econometrics and financial analysis. On the one hand, a logarithmic transformation can be useful due to its variance stabilizing properties. On the other hand, this variance stabilizing property could mask the statistical properties that were intended to be modelled. Furthermore, a logarithmic transformation restricts the effects of fundamental influences to an exponential form, while other non-monotonic relationships may be appealing. A general transformation for time series including values equal to or below zero is given by equation (2-2), where price $p$, on day $t$ and hour $h$ is transformed as follows:

$$x_{h,t} = ln\left(p_{h,t} + e\right) \tag{2-2}$$

A positive constant (here Euler's number e = 2,718...) is added to receive values greater or equal to one. If the original time series is highly non-normal this transformation can reduce skewness and kurtosis of the price distribution. In some electricity markets prices can get negative. Eq. (2-3) suggests an adjusted log-transformation for this case.

$$x_{h,t} = \frac{p_{h,t}}{|p_{h,t}|} ln\left(|p_{h,t}| + 1\right), for\ p\ \neq 0 \tag{2-3}$$

(ii) In the second step, deterministic influences on hourly electricity prices are modelled to isolate the stochastic part of the electricity price formation. The *first* information included in the regression model is the day-ahead price information of relevant commodities. Instead of directly using coal, gas and emission prices we calculate variable cost $c_t^{fuel}$ in order to avoid multicollinearity. The variable costs of a power plant include the fuel prices $(p_t^{fuel})$ and emission certificate prices $\left(p_t^{CO_2}\right)$. By assuming typical plants efficiencies $(\eta^{fuel})$ and emission intensities $(\nu_{CO}^{fuel})$ the following equation delivers the typical variable costs of coal $(c_t^{coal})$ and gas $(c_t^{gas})$ fired power plants. The *second* information included in the regression model is the day-ahead expected residual load $(R_{h,t})$ as the difference between load $(L_{h,t})$ and solar infeed $(S_{h,t})$. Additionally, the expected production of wind $(W_{h,t})$ is included. Wind infeed is introduced as separate variable because it exhibits no daily cycles like load and solar.[3] The residual load data and the wind infeed already captures the seasonal effects, so that it becomes redundant to apply a seasonal function or introduce other dummy variables for weekly patterns or holidays and special days (e. g. Christmas).

$$c_t^{fuel} = \frac{p_t^{fuel} + p_t^{CO_2} \cdot v_{CO2}^{fuel}}{\eta^{fuel}} \tag{2-4}$$

For the $H = 24$ price series, we consider $H$ multiple regressions of the form shown in Eq. (2-4).

$$x_{h,t} = \beta_{h,0} + \beta_{h,1} c_t^{coal} + \beta_{h,2} c_t^{gas} + \beta_{h,3}(L_{h,t} - S_{h,t}) + \beta_{h,4} W_{h,t} + \varepsilon_{h,t} \tag{2-5}$$

The estimation of regression coefficients $\beta_{h,0-4}$ is done with Ordinary Least Squares method (OLS) where $\varepsilon_{h,t}$ is the notation for the residuals. OLS delivers robust results in presence of non-normal errors even if it is not the most efficient estimation procedure.
(iii) In the third step, the residuals $\varepsilon_{h,t}$ are transformed by mapping their empirical distribution to a normal distribution.[4] More precisely, the empirical cumulative distribution function (ECDF) of the residuals is determined. As such, each residual $\varepsilon_{h,t}$ corresponds to a particular quantile of the ECDF. Subsequently, the associated quantile of the normal distribution $u_{t,h}$ is determined by using the inverse of the cumulative distribution function $\Phi$ of the standard normal distribution. The transformation ($T_h$) is represented by:

$$T_h : \varepsilon_{h,t} \longmapsto u_{h,t} = \Phi^{-1}\left(ECDF_h(\varepsilon_{h,t})\right) \tag{2-6}$$

(iv) In the fourth step, the normally distributed residuals $u_{h,t} = T_h(\varepsilon_{h,t})$ for different hours are compared and common effects are identified by performing a factor decomposition based on a PCA. Doing so, we first evaluate the correlations between the standardized errors of every hour. Second, the eigenvalue-eigenvector-problem is solved for the resulting correlation matrix $\boldsymbol{C}$.[5] Since $\boldsymbol{C}$ is by construction a symmetric matrix, there are $H = 24$ real-valued (not necessarily distinct) eigenvalues $\lambda_i$, which are obtained as solutions of the problem:

[3] Solar infeed is not considered separately due to its non-linear form and the high number of zeros in the time series.
[4] Achieving normality is also necessary to be allowed to test ARMA specifications.
[5] Since all dependencies in the correlation matrix are measured in the same unit, it is possible to use the covariance matrix and avoid the division of the covariance to receive the correlations. Yet, in contrast to Zugno et al. (2013) we use the correlation matrix since we do not want to isolate the strongest effect and include all variation in the dataset for the calibration of the stochastic processes.

$$\boldsymbol{Cv_i} = \lambda_i \boldsymbol{v_i}, \text{with } i \in \{1, \dots, I\} \tag{2-7}$$

Thereby the eigenvectors $\boldsymbol{v_i}$ are by definition different from the zero vector. The index $i$ identifies a particular factor and the total number of factors is given by $I = H = 24$. The corresponding eigenvalues $\lambda_i$ represent the strength of an influence factor. The higher the eigenvalue, the more influence the factor has on the total variance of the $u_{h,t}$. The components of the eigenvector $\boldsymbol{v_i}$ can be interpreted as loadings on factor $f_{i,t}$. In matrix notation the relationship between error terms $u_{h,t}$ and factors $f_{i,t}$ is the following:

$$\mathbf{U} = \mathbf{FLV} \tag{2-8}$$

Matrices $\boldsymbol{U}$ and $\boldsymbol{F}$ include all error terms $u_{h,t}$ and all factors $f_{i,t}$. Matrix $\boldsymbol{V}$ contains the orthonormal eigenvectors $\boldsymbol{v_i}$ and the diagonal matrix $\boldsymbol{L}$ the square root of the eigenvalues $\lambda_i$.

$$\mathbf{U} = \begin{pmatrix} u_{1,1} & u_{2,1} \cdots & u_{H,1} \\ u_{1,2} & u_{2,2} \cdots & u_{H,2} \\ \vdots & \vdots \ddots & \vdots \\ u_{1,T} & u_{2,T} \cdots & u_{H,T} \end{pmatrix} \qquad \mathbf{F} = \begin{pmatrix} f_{1,1} & f_{2,1} \cdots & f_{I,1} \\ f_{1,2} & f_{2,2} \cdots & f_{I,2} \\ \vdots & \vdots \ddots & \vdots \\ f_{1,T} & f_{2,T} \cdots & f_{I,T} \end{pmatrix}$$

$$\mathbf{V} = (\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_{24}})^T \qquad \mathbf{L} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \dots & 0 \\ 0 & \sqrt{\lambda_2} \dots & 0 \\ \vdots & \vdots \ddots & \vdots \\ 0 & 0 \dots & \sqrt{\lambda_I} \end{pmatrix} \tag{2-9}$$

(v) In the fifth step we analyze the stochasticity of the factor $f_{i,t}$. To model the stochasticity, different stochastic processes may be considered (see chapter 1). The most sophisticated process for the stochastic component in this paper will be an ARMA(1,1)-GARCH(1,1)[6] specification to model lagged effects in price level and price volatility.[7] An ARMA (1,1)-GARCH(1,1) specification the factor $f_{i,t}$ is dependent on its own lagged value the period before and the error $\omega_{i,t}$ of the current and $\omega_{i,t-1}$ of the previous period (2-10), while the conditional variance follows itself an autoregressive process (2-11).

$$f_{i,t} = \mu_i + \alpha_{i,1} f_{i,t-1} + \alpha_{i,2}\, \omega_{i,t-1} + \omega_{i,t} \; with \; \omega_{i,t} \sim \mathrm{N}(0, \sigma_{i,t}) \tag{2-10}$$

[6] ARMA stands for Autoregressive Moving Average and GARCH for Generalized Autoregressive Conditional Heteroskedasticity, respectively.
[7] In order to prevent over-fitting, the lag lengths of the ARMA-GARCH specifications were determined based on standard information criteria (AIC and BIC) and significance tests.

$$\sigma_{i,t}^2 = \gamma_{i,0} + \gamma_{i,1}\, \sigma_{i,t-1}^2 + \gamma_{i,2}\, w_{i,t-1}^2 \tag{2-11}$$

For the estimation of the parameters $\alpha_{i,1}, \alpha_{i,2}$ and $\gamma_{i,0}$, $\gamma_{i,1}$, $\gamma_{i,2}$, the maximum likelihood method is used, determining the parameters where the probability of the observation of the empirical value of $f_{i,t}$ is maximized. For the parameterization of the model, no outliers are removed from the time series.[8]

(vi) The sixth step is the selection of the rolling window in order to estimate and simulate prices. Note that for the simulation of the price distribution, steps (i) to (v) have to be performed in inverse order. The basic idea of rolling windows is to re-estimate the parameters after each window to update the information basis. If the rolling window length is fixed, older information is falling out of the window as time goes by, so that there is no increasing memory in the future. Fixing the rolling window length is reasonable because of the dynamics in the electricity markets over the past years, the resulting structural changes in price formation and distributional characteristics. This is consistent with previous findings that longer memory arises in electricity markets dominated by hydro rather than thermal generation (Karakatsani and Bunn 2008) and to avoid over estimation (Catalao et al. 2007). Another option of modifying the rolling window techniques is to give more weights to certain observations, e.g. the latest observations (forgetting factor). So-called time adaptive models where already successfully tested by Jónsson et al. 2013.

## 2.3 Evaluation of forecast quality

To evaluate the point forecast quality, common statistical measures like the Mean Absolute Error (MAE) or the Root Mean Square Error (RMSE) may be used. We assess the point forecast quality using the MAE and the RMSE, as both measures are robust to adverse effects of prices close to zero (contrary to the MAPE) and the RMSE in contrast to the MAE puts more weight on larger errors (cf. (Misiorek et al. 2006)). (Weron 2014) as well as Hyndman and Koehler (2006) highlight the importance of scaled error measures and suggest to normalize the error, i.e. by the average price within the evaluation interval of one week, known as Weekly-weighted Mean Absolute Error (WMAE) or Mean Weekly Error (MWE).

[8] Outliers are not removed in order not to distort the estimation results or mask the statistical properties. Influences on estimation results, based on different outlier treatment, is discussed in Trueck et al. (2007). Given that the residuals are transformed to a normal distribution in step (iii), the impact of outliers on the subsequent steps is anyhow reduced.

$$MWE = \frac{1}{\bar{p}_{168}} \cdot \frac{\sum_{h=1}^{168} |p_h - \hat{p}_h|}{168} \tag{2-12}$$

If every hour of a day is modelled separately, it is important to avoid unrealistic jagged forms of individual price paths over the 24 hours of a day.[9] To measure the smoothness of the simulated price paths, we suggest to investigate the spot price variation of every day $t$ until the end of the simulation period $T$ and over each simulation path $n$ of the total number of simulations $N$, see equation (2-13). We therefore define the smoothness indicator $SI$:

$$SI = \frac{\sum_{n=1}^{N} \frac{\sum_{t=1}^{T} \sum_{h=1}^{23} |p_{t,h+1,n} - p_{t,h,n}|}{T}}{N} \tag{2-13}$$

The point forecast error measures do not contain information about the distance between the observations and individual simulation paths and fail to provide information on the quality of probabilistic forecasts. Among others Christoffersen and Diebold (2000) propose to evaluate interval forecasts by calculating the model-dependent prediction interval for the next observation. The nominal coverage is compared to the empirical coverage for different significance levels. The quality assessment of distribution forecasts require the assessment of all interval forecasts, which Diebold et al. (1998) consider a daunting task. Additionally, Pinson et al. (2007) maintain that the evaluation of the empirical coverage of any given interval is not sufficient, since it would be necessary to establish that both quantiles comprising the interval are unbiased.

To evaluate the distribution forecast quality, we investigate the complete distribution forecast rather than a restricted number of quantiles of said forecast. Let $G_{Y_t|\Omega_t} = Pr(Y_t \leq y|\Omega_t)$ denote the conditional distribution of a random variable $Y_t$ given some information set $\Omega_t$ and let $\{G_{Y_t|\Omega_t}\}_{t=1}^{\tau}$ denote the sequence of conditional distributions. The corresponding sequence of one-step-ahead distribution forecasts is denoted by $\{F_{Y_t|\Omega_t}\}_{t=1}^{\tau}$, where $F$ may be of parametric or non-parametric form. If a distribution forecast coincides with the true underlying distribution, it is said to be *calibrated*. We thus test for calibration of the distribution forecast sequence. The null and alternative hypotheses of interest are:

$$\begin{aligned} H_O&: \{G_{Y_t|\Omega_t}\}_{t=1}^{\tau} = \{F_{Y_t|\Omega_t}\}_{t=1}^{\tau}, \\ H_A&: \{G_{Y_t|\Omega_t}\}_{t=1}^{\tau} \neq \{F_{Y_t|\Omega_t}\}_{t=1}^{\tau}. \end{aligned} \tag{2-14}$$

[9] This requirement is particularly important if the price simulations are to be used to evaluate the value of storage-type options.

Under the null hypothesis, the sequence of probability integral transforms (PITs)[10], $\{F_{Y_t|\Omega_t}(Y_t)\}_{t=1}^{\tau}$, is uniformly distributed on [0,1] and independent, given that $\Omega_t$ contains all relevant information. Diebold et al. (1998), who prove the preceding result, propose to test the null hypothesis using a graphical framework, where histograms and correlograms are used to assess uniformity and independence, respectively. Yet, Corradi and Swanson (2006) show that the PITs are still uniformly distributed on [0,1], while the independence result cannot be upheld under the null hypothesis when $\Omega_t$ does not contain all relevant information. Consequently, formal tests for uniformity of the PITs have to account for potential autocorrelation and classic distribution tests, which rely on i.i.d. observations, cannot be applied. Various autocorrelation-robust distribution tests have been suggested in the literature. Yet, neither of these has been specifically designed for assessment of calibration and consequently all suffer major shortcomings in the present context, as identified by Knüppel (2015). The latter introduces a specific calibration test based on raw moments of the PITs, that is robust to dynamic misspecification under the null hypothesis, has power against a variety of distribution misspecifications and is based on standard critical values. The probabilistic forecasting literature establishes the uniformity of the PITs as a necessary but not sufficient condition for ideal distribution forecasts. Since the independence of the PITs is rarely achievable in pure empirical settings, *calibration* of a distribution forecast is considered achieved when the PITs are uniformly distributed. Consequently, further assessment criteria have to be considered to distinguish competing distribution forecasts that fulfil the necessary condition of uniformity of the PITs. Gneiting et al. (2007) propose an evaluation paradigm of maximizing *sharpness subject to calibration*. Sharpness constitutes a property of the distribution forecast only and refers to its concentration. It can be assessed using the average width of central prediction intervals with prespecified nominal coverage or through so-called sharpness diagrams. Additionally, scoring rules such as the Continuous Ranked Probability Score (CRPS) can be used to assess *calibration and sharpness simultaneously*. Competing distribution forecasts are ranked by comparing averages of the respective CRPS sequences. Yet, said averages may be too close to allow for a reliable decision on differences in forecasting performance. Gneiting and Katzfuss (2014) present a Diebold-Mariano-type test based on the CRPS sequences, which allows testing the null hypothesis of equal predictive performance and identifies the preferred specification, if the null hypothesis is rejected. To evaluate the distribution forecasts of the presented methodology, we use histograms and the formal test by Knüppel (2015) to assess calibration and the CRPS with the associated Diebold-Mariano-type test to select the superior distribution forecast among the forecasts that fulfil the necessary condition of uniformity of the PITs.

[10] The Probability Integral Transform is also known in the literature as Rosenblatt (1952) transformation.

# 3 Results

## 3.1 Data

In the application, we consider the results for hourly electricity spot prices in Germany for the year 2015.[11] Table 1 summarizes the distributional parameters of the 24 individual hours. Figure 10 in the appendix visualizes the shape and causal dependencies of the considered panel data.

Table 1: Descriptive Statistics of Hourly Day-ahead Prices 2015

| *Hour* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Min. | -19.98 | -23.06 | -41.74 | -31.41 | -46.97 | -13.48 | -25.02 | -16.94 | -6.86 | -9.11 | -10.31 | -5.39 |
| Max. | 51.68 | 40.83 | 38.90 | 34.92 | 35.32 | 38.97 | 49.95 | 66.14 | 71.92 | 69.68 | 66.23 | 62.97 |
| Mean. | 25.17 | 23.27 | 21.90 | 21.29 | 21.69 | 23.79 | 30.08 | 36.89 | 38.87 | 36.99 | 34.93 | 34.21 |
| SD | 9.13 | 9.33 | 9.97 | 9.20 | 9.19 | 8.36 | 11.46 | 14.14 | 13.86 | 12.37 | 11.84 | 11.57 |
| Skewness | -1.79 | -1.94 | -2.24 | -1.82 | -2.21 | -1.43 | -1.11 | -0.67 | -0.42 | -0.24 | -0.15 | -0.03 |
| Kurtosis | 8.46 | 8.03 | 10.32 | 7.75 | 12.44 | 5.71 | 5.06 | 3.39 | 3.05 | 3.30 | 3.32 | 3.13 |
| No. neg. | 10 | 15 | 19 | 14 | 14 | 9 | 9 | 6 | 2 | 1 | 2 | 2 |
| *Hour* | *13* | *14* | *15* | *16* | *17* | *18* | *19* | *20* | *21* | *22* | *23* | *24* |
| Min. | -11.05 | -65.06 | -79.94 | -65.02 | -19.11 | 5.24 | 11.44 | 10.55 | 8.49 | 4.04 | 7.34 | -7.50 |
| Max. | 60.00 | 59.59 | 60.04 | 65.05 | 70.82 | 99.77 | 84.94 | 98.05 | 65.05 | 59.94 | 61.95 | 52.49 |
| Mean. | 31.58 | 30.10 | 29.41 | 30.51 | 32.24 | 37.80 | 41.63 | 42.53 | 38.69 | 34.90 | 33.23 | 27.35 |
| SD | 11.11 | 12.32 | 12.93 | 12.70 | 11.81 | 12.38 | 12.03 | 11.39 | 9.05 | 8.31 | 8.03 | 7.96 |
| Skewness | -0.14 | -1.33 | -1.88 | -1.34 | -0.29 | 0.38 | 0.42 | 0.44 | -0.25 | -0.43 | -0.31 | -1.08 |
| Kurtosis | 3.64 | 12.47 | 17.26 | 11.70 | 4.07 | 4.43 | 3.70 | 4.40 | 3.48 | 3.94 | 4.16 | 5.87 |
| No. neg. | 3 | 4 | 5 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |

Generally, all variables of the deterministic part of the regression have to be forecasted in order to forecast and simulate electricity prices. In the application, we forecast the next 24 hours and use publically available day-ahead forecast information of the fundamental inputs or use short-term myopic forecasts (notably for the fuel prices) without doing any prior simulations of the inputs. Table 2 summarizes the data sources.[12]

[11] To check validity and robustness we applied the methodology and evaluation also for the year 2014.
[12] To calculate the typical variable costs of power plants we use day-ahead gas prices from the TTF and for coal prices the API#2 font year notation. We approximate the coal plant's efficiency to be 0.40 and emission intensity to be 0.3 t/MWh. For the gas power plant we assume 0.50 efficiency and 0.2 emission intensity.

Table 2: Data Set[13]

| *Data* | *Source* | *Product* |
|---|---|---|
| Load | ENTSO-E | - |
| Spot prices | EPEX-Spot | Info-User (EOD) |
| Coal Price | Energate.de | API#2 Front year |
| $CO_2$ | EEX | EU-$CO_2$-Emission Allowances |
| Gas price | Spectron | TTF-Day-Ahead |
| Wind forecast and realization | EEX Transparency data service | Info-Vendor |
| Solar forecast and realization | EEX Transparency data service | Info-Vendor |

A commonly used benchmark is the so called Naïve forecast as in Conejo et al. (2005). The Naïve forecast are the 24 prices of the last similar day (e. g. last Saturday for next Saturday). A forecast method that is unable to outperform the Naïve Benchmark is not sufficient. Another common benchmark for electricity price forecasting is a simple autoregressive (AR) process. Table 3 summarizes the various specifications considered in the present study.

Table 3: Model specifications (The 'x' in each row indicates which model choices are combined)

| | *AR(1)* | *AR(2)* | *ARMA(1,1)* | *GARCH(1,1)* | *PCA* |
|---|---|---|---|---|---|
| **Model 1** | x | | | | |
| **Model 2** | | | x | x | |
| **Model 3** | | x | | | |
| **Model 4** | x | | | | x |
| **Model 5** | | | x | x | x |
| **Model 6** | | x | | | x |

## 3.2 Estimation results

The first step in the presented estimation procedure is the transformation of the data (see Chapter 2). As indicated by Table 1, the moments of the untransformed prices 2015 are already close to those of a normal distribution. Thus, for our application it is beneficial not to apply the transformation. Since this is not generally the case and due to the mentioned properties of a log-transformation, we recommend a careful check before the application of step (i). The results from the second step are the parameter estimates for the deterministic influences. The values for every single hour are summarized in Table 4

[13] Treatment of time shift days: in spring (day = 23 hours), hour three was added as missing value (NV). In autumn (day = 25 hours), hour 3 b was deleted from the sample. Missing values were filled with their previous value (either 1 hour or 24 hours before e. g. in case of solar power production)

for the initial rolling window (01.01.2013 until 31.12.2014). Over time, small changes for the coefficient estimates occur. Autocorrelation of residuals is accounted for in later steps.

Table 4: Estimation results of initial rolling window

| ***Hour*** | | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** | ***6*** | ***7*** | ***8*** |
|---|---|---|---|---|---|---|---|---|---|
| | $\beta_{h,0}$ | -9.79* | -19.25* | -29.00* | -38.86* | -43.28* | -50.92* | -65.69* | -66.26* |
| Coal | $\beta_{h,1}$ | 0.69* | 0.62* | 0.85* | 0.99* | 1.24* | 1.68* | 1.32* | 0.58** |
| Gas | $\beta_{h,2}$ | 0.14* | 0.16* | 0.14* | 0.17* | 0.19* | 0.22* | 0.40* | 0.63* |
| Load | $\beta_{h,3}$ | 0.46* | 0.68* | 0.78* | 0.85* | 0.75* | 0.61* | 0.96* | 1.20* |
| Wind | $\beta_{h,4}$ | -1.37* | -1.52* | -1.64* | -1.61* | -1.47* | -1.27* | -1.28* | -1.54* |
| $R^2$ | % | 70.70 | 71.97 | 68.81 | 67.22 | 65.73 | 66.68 | 75.22 | 78.67 |
| ***Hour*** | | ***9*** | ***10*** | ***11*** | ***12*** | ***13*** | ***14*** | ***15*** | ***16*** |
| | $\beta_{h,0}$ | -39.66* | -39.49* | -37.58* | -19.71* | -23.41* | -27.72* | -32.68* | -42.21* |
| Coal | $\beta_{h,1}$ | -0.24 | 0.08 | 0.24 | -0.25 | 0.07 | 0.02 | -0.01 | 0.23 |
| Gas | $\beta_{h,2}$ | 0.65* | 0.63* | 0.55* | 0.50* | 0.43* | 0.42* | 0.41* | 0.43* |
| Load | $\beta_{h,3}$ | 1.13* | 1.01* | 0.96* | 0.92* | 0.87* | 0.96* | 1.05* | 1.07* |
| Wind | $\beta_{h,4}$ | -1.65* | -1.69* | -1.63* | -1.61* | -1.48* | -1.53* | -1.52* | -1.44* |
| $R^2$ | % | 72.31 | 73.62 | 73.26 | 67.81 | 70.80 | 73.09 | 70.65 | 71.63 |
| ***Hour*** | | ***17*** | ***18*** | ***19*** | ***20*** | ***21*** | ***22*** | ***23*** | ***24*** |
| | $\beta_{h,0}$ | -60.49* | -85.45* | -110.18* | -72.43* | -44.08* | -13.10* | 13.55* | 10.76** |
| Coal | $\beta_{h,1}$ | 0.75* | 1.12* | 2.17* | 1.26* | 1.03* | 0.58* | 0.18 | 0.48** |
| Gas | $\beta_{h,2}$ | 0.51* | 0.72* | 0.80* | 0.73* | 0.56* | 0.39* | 0.30* | 0.14* |
| Load | $\beta_{h,3}$ | 1.06* | 1.17* | 1.16* | 0.98* | 0.70* | 0.45* | 0.19* | 0.13* |
| Wind | $\beta_{h,4}$ | -1.35* | -1.41* | -1.59* | -1.55* | -1.44* | -1.38* | -1.26* | -1.08* |
| $R^2$ | % | 76.32 | 72.18 | 72.55 | 62.51 | 61.14 | 55.80 | 56.11 | 45.92 |

Significance is computed using standard errors obtained through the Newey-West procedure. Significance at 5% level is labeled with ** and at 1% level with *.

The graphical representation of the quantile mapping performed in step (iii) is the Q-Q-Plot (Figure 1).

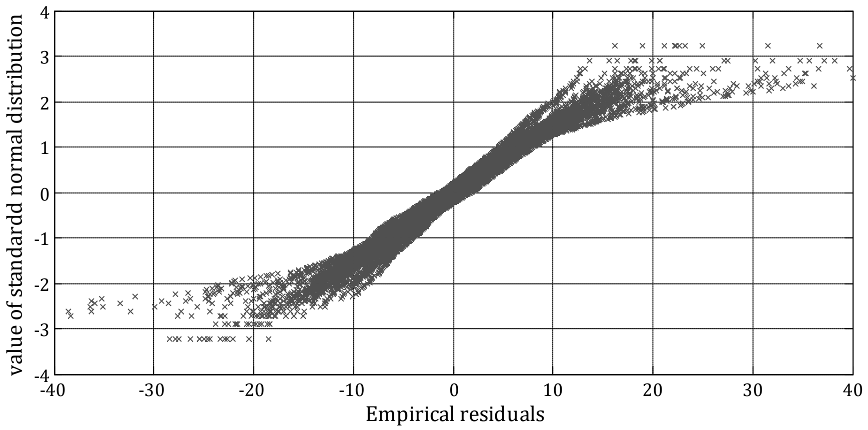


Figure 1: Q-Q-Plot corresponding to the transformation function $T_h$ for $\varepsilon_{h,t}$ into $u_{h,t}$.

In step (iv) the PCA is performed. The resulting eigenvalues λ are shown in Figure 2. The higher the eigenvalue, the higher is the contribution of the respective factor to explaining the variance of $u_{h,t}$. The eigenvalues decrease rapidly and the first eigenvalue explains the major part of the observed variance. The eigenvalues do not change significantly for the different windows when rolling over time.

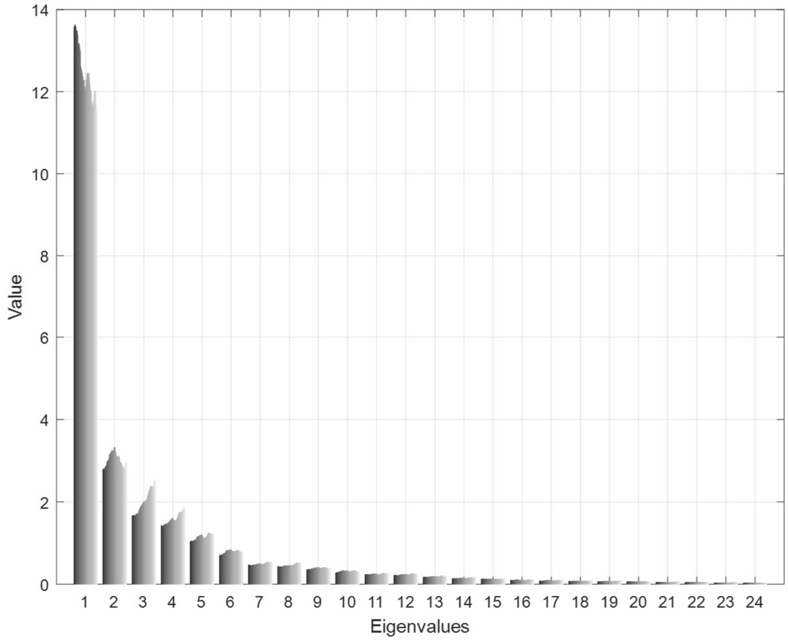


Figure 2: Eigenvalues λ of the correlation matrix C for the rolling windows of 2015[14]

A potential weakness of the PCA is the lack of interpretation of the obtained factors. Sometimes it is difficult or impossible to specifically identify the individual factors or find an economic interpretation. Figure 3 shows the first six eigenvectors for all rolling windows.

Factor one exerts a rather similar influence over all hours and can be interpreted as the base price component. Furthermore, factor two has high positive loadings on off-peak hours and negative loadings on peak hours. It is consequently interpreted as off-peak – peak spread. Thus, factor two also captures the effect of the sun peak (approx. 11 am to 4 pm). The increased variability of factor loadings during the sun-peak hours is due to the seasonality of solar infeed. Factor three to five also exhibit a somewhat increased variability. They all describe the relative strength of some evening hours against some morning hours and may be summarized as shape components. For factor six, we again establish a rather stable pattern. It exerts strong influence during the morning and evening peaks and may thus be interpreted as ramping component. The remaining factors are not as readily interpreted fundamentally. However, they do not influence the shape of

[14] The figure shows bar plots of the 24 eigenvalues of the correlation matrix C for the rolling windows of 2015. Different shades of grey represent estimates of a particular eigenvalue for different rolling windows.

hourly prices as strongly, as indicated by the magnitude of the associated eigenvalues in Figure 3.

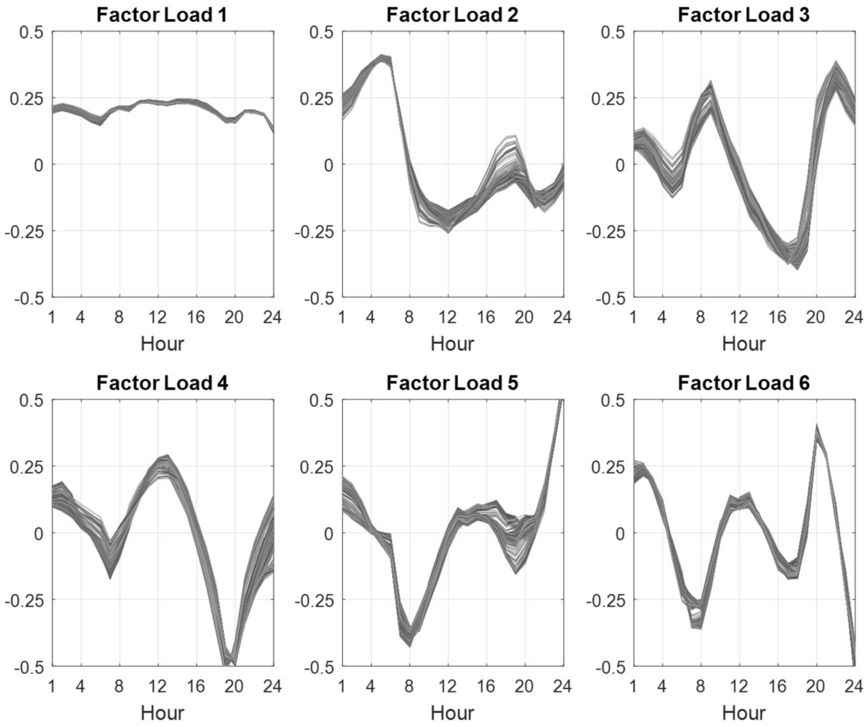


Figure 3: Factor loadings of the first six factors for all rolling windows in 2015

### 3.3 Forecast results

Following the testing procedure outlined in chapter 2.3, a first *qualitative judgement* is given by an eyeball investigation of the simulation results (see Figure 4).

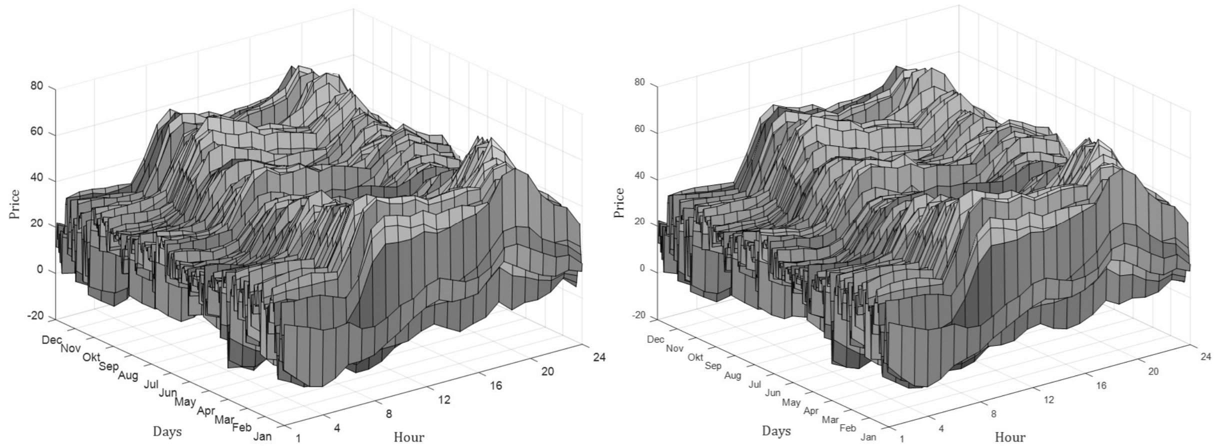


Figure 4: Surface of mean simulation results for 2015: model 3 (left) and model 6 (right)

The eyeball investigation for the mean across all Monte Carlo simulations shows that the main characteristics of electricity prices, e.g. daily cycles, are captured and no systematic errors occur. The comparison of the specification with PCA (model 6) and without PCA (model 3) uncovers a tendency of the PCA to smoothen the daily point forecast curves.

### 3.4 Evaluation of forecast quality

Descriptive statistics and forecast error measures for point forecasts, calculated as averages over the simulation paths, are summarized in Table 5 and Table 6. The various specifications of the proposed methodology work accurately and exhibit statistical behavior that is close to the actual data. Yet, one has to note that extreme prices are not well replicated and that all specifications underestimate the kurtosis of the actual data. Since the point forecasts are calculated as mean over all Monte Carlo paths, extreme events and substantial price changes are rounded off.

Table 5: Descriptive statistics of point forecasts for 2015

| | ***Data*** | ***M1*** | ***M2*** | ***M3*** | ***M4*** | ***M5*** | ***M6*** | ***Naïve*** |
|---|---|---|---|---|---|---|---|---|
| **Min.** | -79.94 | -15.62 | -14.54 | -10.78 | -11.70 | -9.93 | -9.44 | -41.74 |
| **Max.** | 99.77 | 78.85 | 72.79 | 72.26 | 74.06 | 70.01 | 70.55 | 99.77 |
| **Mean** | 31.63 | 30.97 | 31.10 | 31.21 | 31.20 | 31.32 | 31.38 | 32.58 |
| **SD** | 12.67 | 12.86 | 12.75 | 12.81 | 12.88 | 12.82 | 12.86 | 11.72 |
| **Skewness** | -0.31 | -0.12 | -0.11 | -0.10 | -0.13 | -0.12 | -0.12 | 0.16 |
| **Kurtosis** | 5.76 | 3.13 | 3.07 | 3.06 | 3.07 | 3.03 | 3.03 | 4.54 |

The point forecast error measures show the proposed methodology to clearly outperform the naïve benchmark. The Mean Error (ME) is closer to zero for all considered specifications and all Mean Absolute Errors (MAE) are at least 2.70 EUR/MWh below the MAE of the naïve benchmark. Furthermore, a much higher share of total price variation is explained by the presented methodology as indicated by the much higher R-Squared. The considered specifications show little variation in point forecasting although model 1 and model 4 appear to lag somewhat behind the results of the other specifications.

Table 6: Error measures of point forecasts for 2015

| | ***M1*** | ***M2*** | ***M3*** | ***M4*** | ***M5*** | ***M6*** | ***Naïve*** |
|---|---|---|---|---|---|---|---|
| Mean Error (ME) | 0.65 | 0.52 | 0.42 | 0.42 | 0.30 | 0.24 | -0.96 |
| Mean Absolute Error (MAE) | 4.27 | 4.06 | 4.05 | 4.25 | 4.09 | 4.08 | 7.03 |
| Root Mean Square Error | 5.74 | 5.48 | 5.47 | 5.71 | 5.52 | 5.50 | 9.84 |
| R-Squared ($R^2$) | 79.47 | 81.25 | 81.38 | 79.66 | 81.00 | 81.15 | 39.70 |

The point forecast error measures for individual hours (Figure 5) support the preceding conclusion. Again, the naïve benchmark is outperformed by all other specifications, which do not diverge significantly for individual hours. The MWEs indicate equal point forecast performance (see Table 12 in the appendix).

It is important to adequately account for cross-correlation between the hours under the considered panel approach in order to generate smooth price paths. Table 7 summarizes the introduced smoothness indicator for all specifications. The actual price series of 2015 exhibits a smoothness indicator of 77.00. Clearly, accounting for cross-correlation patterns through a PCA substantially improves the smoothness of the simulated price paths (models 4 to 6). Additionally, allowing for conditional heteroscedasticity through GARCH-effects (model 5) leads to a smoothness indicator that is nearly identical to that of the actual prices series.

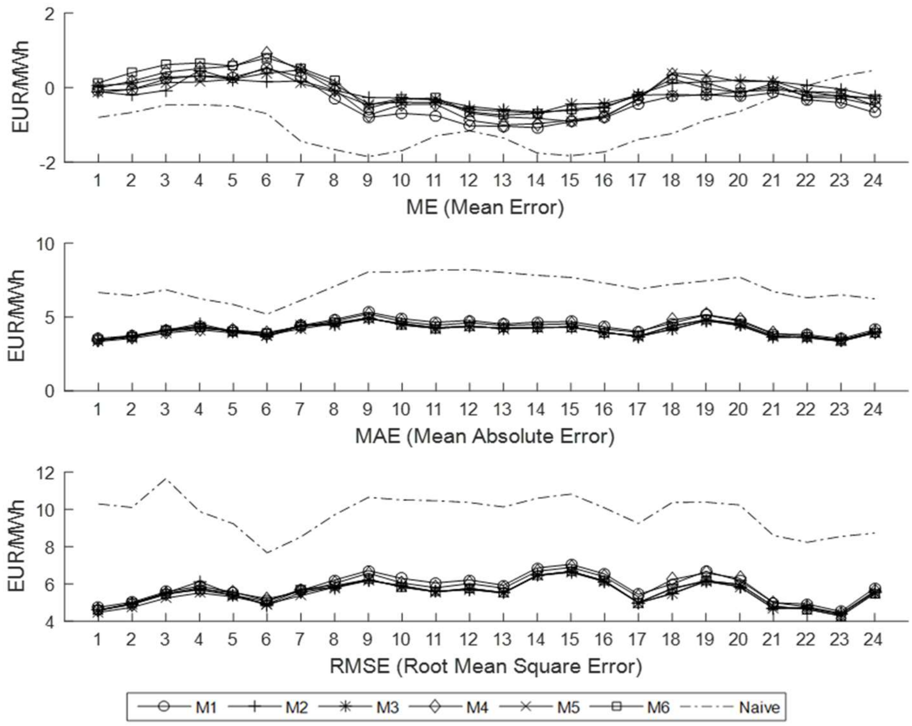


Figure 5: Point forecast errors per hour of 2015

Table 7: Smoothness indicator for 2015

| | *Data* | *M1* | *M2* | *M3* | *M4* | *M5* | *M6* |
|---|---|---|---|---|---|---|---|
| ***Smooth*** | 77.00 | 155.28 | 117.72 | 150.65 | 91.03 | 77.97 | 91.20 |

Table 8 summarizes the empirical coverage for central prediction intervals with nominal coverage of 50, 90, 95 and 99 per cent, respectively.[15]

[15] We have argued in Section 2.3 that the evaluation of interval forecasts is not sufficient to assess the presented probabilistic forecasts and that specific evaluation techniques for distribution forecasts have to be considered. The results of interval evaluation are nevertheless included to assess to what extent they support the conclusions drawn using distribution forecast validation techniques.

Table 8: Interval forecasting results for the whole sample (# Inc. = number of observations in the CI)

| Nominal coverage | | M1 | | M2 | | M3 | | M4 | | M5 | | M6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Inc. | CI-Level | #Inc. | % | #Inc. | % | #Inc. | % | #Inc. | % | #Inc. | % | #Inc. | % |
| 4'380 | **50** | 4'144 | 47 | 3'020 | 34 | 4'235 | 48 | 4'125 | 47 | 3'211 | 37 | 4'183 | 48 |
| 7'884 | **90** | 7'710 | 88 | 6'325 | 72 | 7'769 | 89 | 7'651 | 87 | 6'696 | 76 | 7'715 | 88 |
| 8'322 | **95** | 8'165 | 93 | 6'967 | 80 | 8'206 | 94 | 8'146 | 93 | 7'353 | 84 | 8'180 | 93 |
| 8'672 | **99** | 8'615 | 98 | 7'790 | 84 | 8'601 | 98 | 8'576 | 98 | 8'108 | 93 | 8'571 | 98 |

The results show that the specifications with conditional heteroscedasticity effects (model 2 and model 5) fail to deliver unbiased interval forecasts, as indicated by the substantial deviation between nominal and empirical coverage, which does not bode well for their ability to deliver reliable distribution forecasts. Considering the prediction intervals with 50 and 99 percent nominal coverage for individual hours supports the previous findings (Figure 6). Model 2 and model 5 are unable to deliver reliable prediction intervals. Furthermore, the remaining specifications appear to have problems to reliably forecast the prediction intervals of off-peak hours.

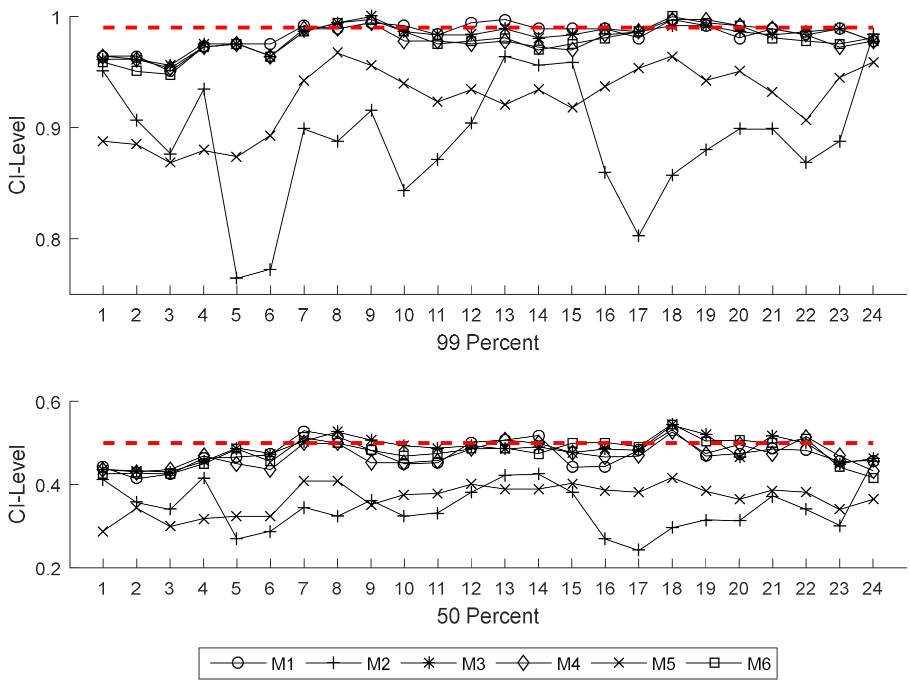


Figure 6: Interval forecasting results per hour of 2015 for 50 and 99 percent nominal coverage

To evaluate the quality of the distribution forecasts directly, we first consider the histograms of the PIT values plotted with 95% Wilson Score Confidence Bands for all hours of 2015 (Figure 7). Following Gneiting et al. (2007), the number of bins has been set to 20. The u-shaped histograms of model 2 and model 5 indicate that these distribution forecasts are too narrow with too many realizations falling in the tails of the forecast distributions. Based on the remaining histograms, we conclude that the required uniformity of PIT values is best achieved by model 3 and model 6, which thus constitute the preferred specifications.

Continuing with the preferred specifications only, the evaluation of the distribution forecasts of model 3 and model 6 for each individual hour allows for a more rigorous assessment of distribution forecast accuracy. They are presented in Figure 8 and Figure 9. We conclude that the uniformity of the PIT values is achieved for model 3, although too many observations fall into the right tail of the distribution for hours 1 to 7. The results for model 6, which extends model 3 by the PCA, are identical.

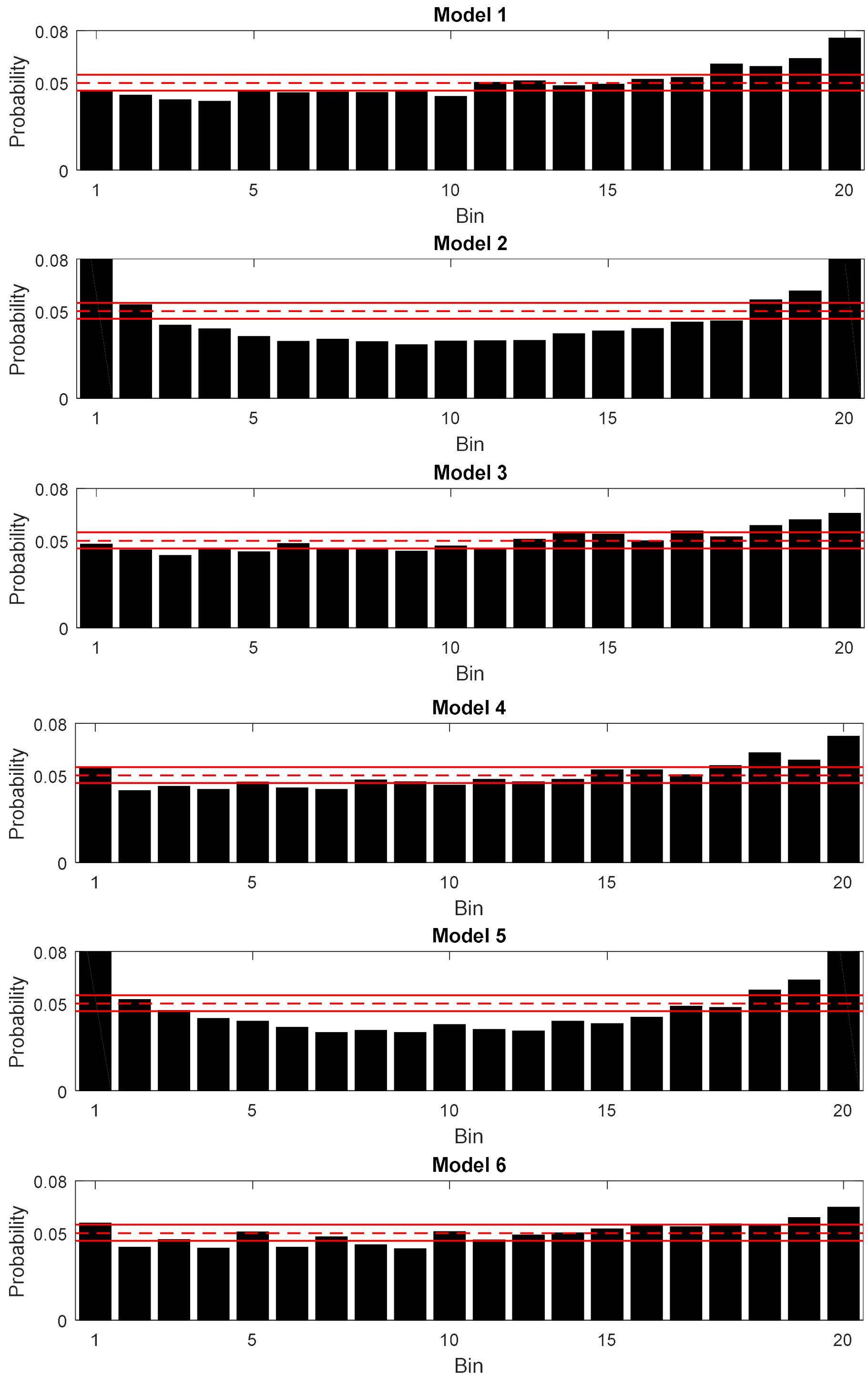


Figure 7: PIT histograms for all 8760 hours of 2015 (with 95% Wilson score confidence bands)

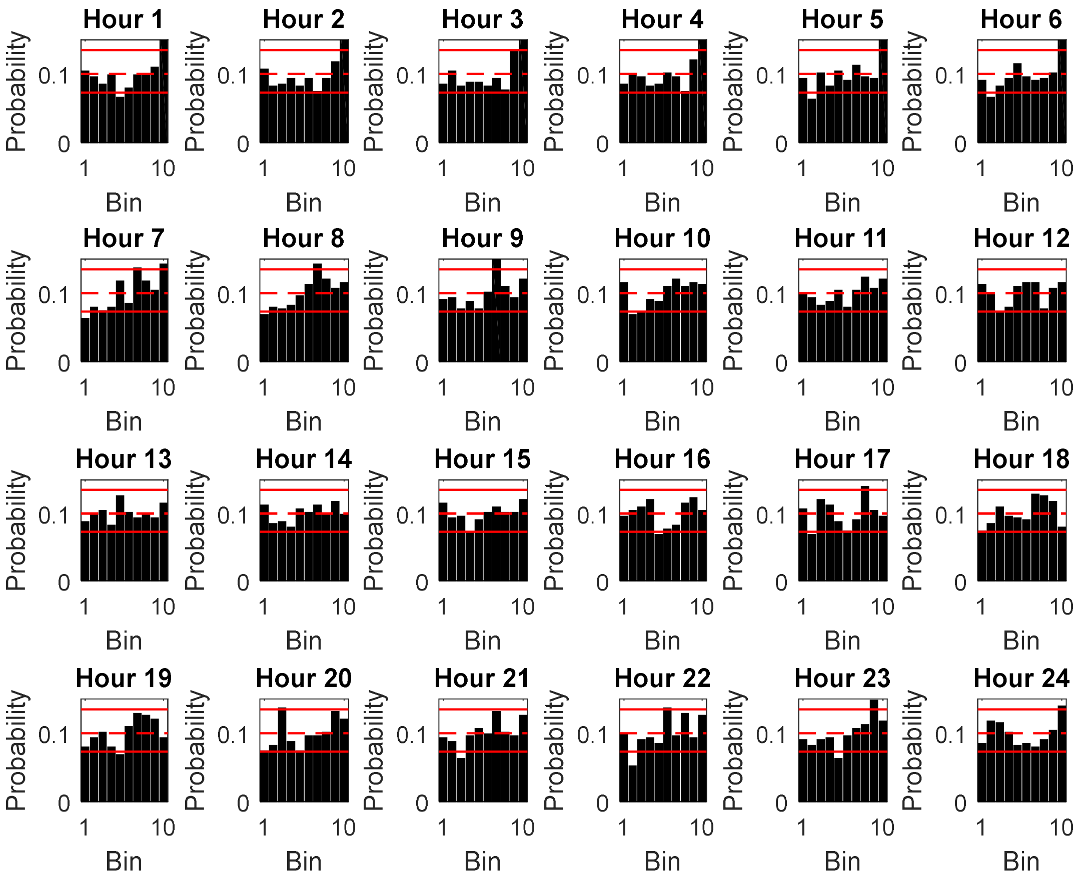


Figure 8:: PIT histograms per hour of 2015 for model 3 (with 95% Wilson score confidence bands)

In addition to the graphical assessment, we consider the formal test for calibration proposed by Knüppel (2015). Table 9 shows the results of the formal test at 1% significance level, where a 1 indicates the failure to reject the null hypothesis of calibration of the distribution forecast sequence for a particular hour. The test results for model 2 and model 5 underscores the forecast deficiencies that have been identified through the graphical assessment with only one hour where the null hypothesis cannot be rejected. Furthermore, the formal test confirms the conclusions of the preceding graphical analysis. We fail to reject the null hypothesis of calibration for 22 and 20 hours of 2015 for model 3 and model 6, respectively. Thus, we conclude that our presented econometric-stochastic approach delivers distribution forecasts that coincide with the actual price distribution in the vast majority of hours. Note that the presented results support the interval forecast evaluation to the extent that the hours for which the null hypothesis of calibration is rejected are mostly among the off-peak hours, which showed the greatest deviation between nominal and empirical coverage.

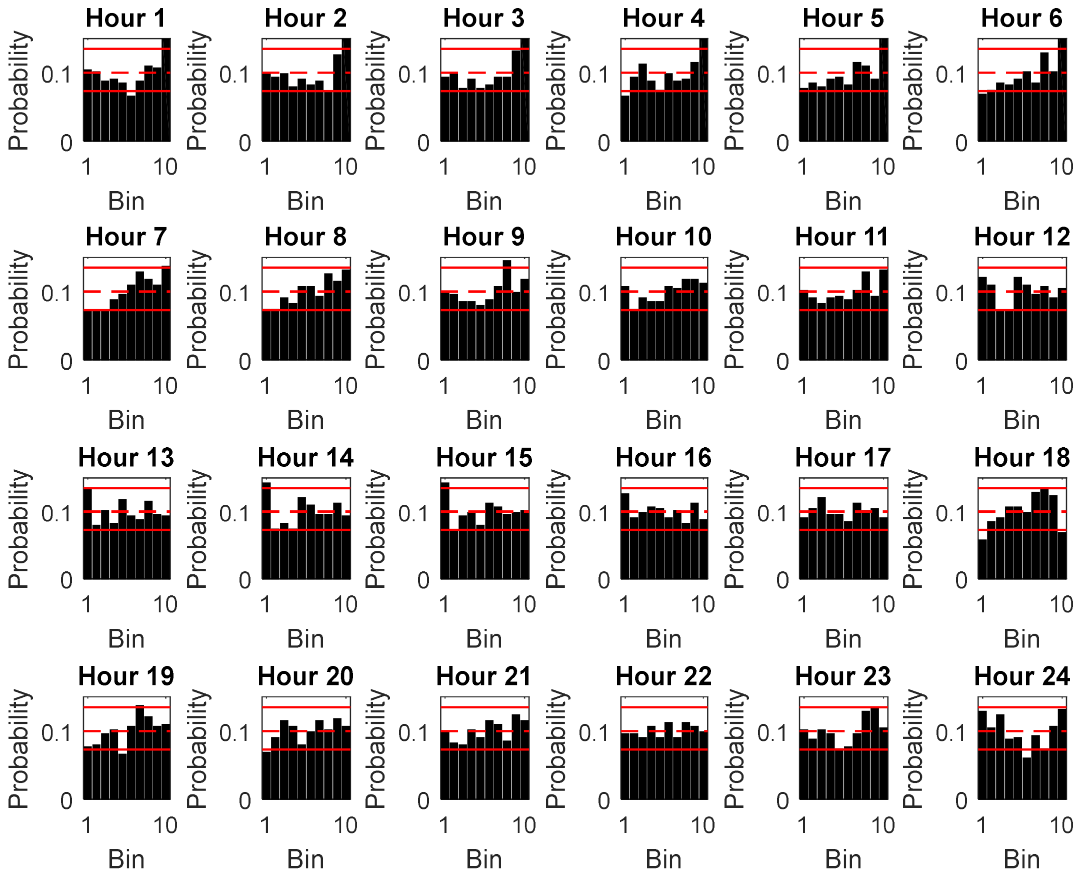


Figure 9: PIT histograms per hour of 2015 for model 6 (with 95% Wilson score confidence bands)

Table 9: Knüppel (2015) test per hour of 2015 (α=1%)

| ***Hour*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** | ***6*** | ***7*** | ***8*** | ***9*** | ***10*** | ***11*** | ***12*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Model 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Model 3 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| Model 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Model 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Model 6 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ***Hour*** | ***13*** | ***14*** | ***15*** | ***16*** | ***17*** | ***18*** | ***19*** | ***20*** | ***21*** | ***22*** | ***23*** | ***24*** |
| Model 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Model 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Model 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Model 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Model 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Model 6 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |

*Failure to reject the null hypothesis of calibration of the distribution forecast sequence is indicated by 1.*

Model 3 and model 6 provide uniformly distributed PIT values and thus meet the necessary condition for calibrated distribution forecasts. We have established that model 6 provides forecasts that are superior with respect to the smoothness measure. To assess whether to prefer model 3 or model 6, we consider the mean CRPS values presented in Table 10. Clearly, the differences between the mean CRPS values for the individual hours of 2015 are small.

Table 10: Mean CRPS Values per Hour of 2015

| ***Hour*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** | ***6*** | ***7*** | ***8*** | ***9*** | ***10*** | ***11*** | ***12*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model 3 | 2.60 | 2.55 | 2.74 | 2.56 | 2.55 | 2.35 | 2.66 | 3.15 | 3.70 | 3.41 | 3.18 | 3.24 |
| Model 6 | 2.55 | 2.50 | 2.70 | 2.53 | 2.58 | 2.44 | 2.60 | 3.20 | 3.76 | 3.40 | 3.21 | 3.28 |
| ***Hour*** | ***13*** | ***14*** | ***15*** | ***16*** | ***17*** | ***18*** | ***19*** | ***20*** | ***21*** | ***22*** | ***23*** | ***24*** |
| Model 3 | 3.01 | 3.14 | 3.19 | 2.90 | 2.73 | 2.94 | 3.37 | 3.24 | 2.77 | 2.75 | 2.46 | 3.08 |
| Model 6 | 3.05 | 3.15 | 3.19 | 2.89 | 2.77 | 3.12 | 3.42 | 3.29 | 2.81 | 2.66 | 2.42 | 3.05 |

We thus use the Diebold-Mariano-type test to assess whether the mean CRPS differentials are statistically significant. Table 11 summarizes the results of the test at 5% and 1% significance level, respectively, and indicates the preferred model in case of non-equal forecast performance. At 5% significance level, there are five instances where the null hypothesis of equal predictive performance is rejected. In three of these five cases the test maintains model 6 to exhibit better predictive performance. Since model 6 additionally provides smoother price paths and is only marginally underperformed by model 3 in terms of calibration, we conclude that model 6 constitutes the preferred specification.

Table 11: Diebold Mariano test of equal predictive performance per hour of 2015

| ***Hour*** | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** | ***6*** | ***7*** | ***8*** | ***9*** | ***10*** | ***11*** | ***12*** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Diebold Mariano test ($\alpha$ = 5%) | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| DM test preferred model ($\alpha$ = 5%) | M6 | M6 | - | - | - | M3 | - | - | - | - | - | - |
| Diebold Mariano test ($\alpha$ = 1%) | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| DM test preferred model ($\alpha$ = 1%) | - | - | - | - | - | M3 | - | - | - | - | - | - |
| ***Hour*** | ***13*** | ***14*** | ***15*** | ***16*** | ***17*** | ***18*** | ***19*** | ***20*** | ***21*** | ***22*** | ***23*** | ***24*** |
| Diebold Mariano test ($\alpha$ = 5%) | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| DM test preferred model ($\alpha$ = 5%) | - | - | - | - | - | M3 | - | - | - | M6 | - | - |
| Diebold Mariano test ($\alpha$ = 1%) | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| DM test preferred model ($\alpha$ = 1%) | - | - | - | - | - | M3 | - | - | - | - | - | - |

*Failure to reject the null hypothesis of equal predictive performance is indicated by 1.*

## 4 Discussion and Conclusion

Forecasting electricity spot prices has to address different issues: (1) Handle the idiosyncratic influences on electricity prices, (2) capture the main characteristics of prices dynamics, (3) deliver not only point forecasts but also distribution forecasts and (4) establish an estimation and simulation procedure that keeps time and effort small.
Therefore, this works presents an approach that improves day-ahead electricity price distribution forecasting by accounting for cross correlation patterns between individual hours. In a stepwise procedure of testing point, interval and distribution forecast accuracy, it is shown that our approach is able to capture the main characteristics of hourly electricity prices and produces reliable point, interval and distribution forecasts for decision support. Among the considered specifications, we find little variation in point forecasting ability. Yet, we argue that an investigation of point forecasts only is not sufficient to consider a particular model to be superior. The same is true for interval forecasts. Rather, an evaluation of distribution forecast ability is required to assess a forecasting model fully. Consequently, only additional tests for calibration and the consideration of the CRPS and associated tests facilitate said evaluation (see section 2.3 and 3.4). The specifications accounting for conditional heteroscedasticity (model 2 and model 5) do not provide reliable distribution forecasts, despite possessing good point forecasts. It is shown that an AR(2) specification without PCA or with PCA (model 3 and model 6) produces smoother simulation paths and reliable distribution forecasts. Our method is flexible in the sense that it can be combined with different point-forecasting approaches for electricity prices and that it can be easily applied to different data exhibiting within-day correlation structures like electricity, e. g. electricity load or heating demand. In combination with a simple regression model, it is appropriate to update the estimation basis using a rolling window because distributional characteristics of energy prices change over time. The length of the rolling window includes a trade-off between robustness of estimation and changes of distributional properties (especially due to the increasing share of renewables). Apart from simple regression models, more sophisticated supply stack modelling approaches are expected to improve the point-forecast quality and eventually the accuracy of the distributional forecasts. Especially a fundamental model that accounts for non-linear dependencies in spot price formation could be considered. Stochastic approaches in general cannot capture fundamental changes, thus detailed fundamental models tend to outperform the pure time series approaches. The combination of the factor decomposition for cross-sectional panels and the quantile mapping improve the forecast ability by smoothing the forecast results and eventually enabling accurate distributional forecasting.

## 5 References

Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets. A review and evaluation. Electrical Power and Energy Systems, 13–22.

Bierbrauer, M., Menn, C., Rachev, S.T., Trueck, S., 2007. Spot and derivative pricing in the EEX power market. Journal of Banking & Finance, 3462–3485.

Catalao, J.P. S., Mariano, S.J. P. S., Mendes, V.M. F., Ferreira L. A. F. M., 2007. Short-term electricity price forecasting in a competitive market. A neural network approach. Electric Power Systems Research, 1297–1304.

Christoffersen, P.F., Diebold, F.X., 2000. How Relevant is Volatility Forecasting for Financial Risk Management? The Review of Economics and Statistics, 12–22.

Conejo, A.J., Contreras, J., Espínola, R., Plazas, M.A., 2005. Forecasting electricity prices for a day-ahead pool-based electric energy market. International Journal of Forecasting, 435–462.

Corradi, V., Swanson, N.R., 2006. Predictive Density and Conditional Confidence Interval Accuracy Tests. Journal of Econometrics, 187–228.

Cuaresma, C.J., Hlouskova, J., Kossmeier, S., Obersteiner, M., 2004. Forecasting electricity spot prices using linear univariate time series models. Applied Energy, 87–106.

Diebold, F.X., Gunther T. A., Tay A. S., 1998. Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 868–883.

Gneiting, T., Balabdaoui, F., Raftery, A., 2007. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69, 243–268 (1 August, 2014).

Gneiting, T., Katzfuss, 2014. Probabilistic Forecasting. Annu. Rev. Stat. App., 125–151. 10.1146/annurev-statistics-062713-085831.

Härdle, W.K., Trueck, S., 2010. The dynamics of hourly electricity prices, Discussion Paper, Berlin.

Huisman, R., Huurman, C., Mahieu, R., 2007. Hourly electricity prices in day-ahead markets. Energy Economics, 240–248.

Huisman, R., Kilic, M., 2013. A History of European Electricity Day-Ahead Prices. Applied Economics, 45, 2683–2693 (1 August, 2014).

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting, 679–688.

Janczura, J., Weron, R., 2010. An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Economics, 1059–1073.

Jónsson, T., Pinson, P., Nielsen, H., Madsen, H., Nielsen, T., 2013. Forecasting Electricity Spot Prices Accounting for Wind Power Predictions. Sustainable Energy, IEEE Transactions on, 4, 210–218.

Jónsson, T., Pinson, P., Madsen, H., Nielsen, H., 2014. Predictive Densities for Day-Ahead Electricity Prices Using Time-Adaptive Quantile Regression. Energies, 7, 5523-5547.

Karakatsani, N.V., Bunn, D.W., 2008. Intra-day and regime-switching dynamics in electricity price formation. Energy Economics, 30, 1776–1797 (1 August, 2014).

Keles, D., Genoese, M., Möst, D., Ortlieb, S., Fichtner, W., 2012. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. Energy Economics, 34, 1012–1032.

Keles, D., Genoese, M., Möst, D., Ortlieb, S., Fichtner, W., 2013. A combined modelling approach for wind power feed-in and electricity spot prices. Energy Policy, 59, 213–225.

Knüppel, M., 2015. Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments. Journal of Business & Economic Statistics, 270–281.

Kosater, P., Mosler, K., 2006. Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices. Applied Energy, 943–958.

Kovacevic, R., Wozabal, D., 2014. A semiparametric model for electricity spot prices. IIE Transactions.

Misiorek, A., Trueck, S., Weron, R., 2006. Point and Interval Forecasting of Spot Electricity Prices. Linear vs. Non-Linear Time Series Models. Nonlinear Analysis of Electricity Prices. Studies in Nonlinear Dynamics & Econometrics.

Nowotarski, J., Weron, R., 2016. Recent advances in electricity price forecasting. A review of probabilistic forecasting, Warschau.

Pape, C., Hagemann, S., Weber, C., 2016. Are Fundamentals Enough? Explaining Price Variations in the German Day-Ahead and Intraday Power Market. Energy Economics, 2016, 376–387. doi:10.1016/j.eneco.2015.12.013.

Pape, C., Woll, O., Weber, C., 2017. Estimating the value of flexibility from real options. On the adequacy of hybrid electricity price models, Essen.

Pinson, P., Chevallier, C., Kariniotakis, G., 2007. Trading Wind Generation From Short-Term Probabilistic Forecasts of Wind Power. IEEE Transactions on Power Systems, 22, 1148–1156.

Rosenblatt, M., 1952. Remarks on a Multivariate Transformation. The Annals of Mathematical Statistic, 470–472.

Trueck, S., Weron, R., Wolff, R.C., 2007. Outlier Treatment and Robust Approaches for Modeling Electricity Spot Prices. Proceedings of the 56th Session of the International Statistical Institute. http://mpra.ub.uni-muenchen.de/4711/.

Ventosa, M., Baíllo, Á., Ramos, A., Rivier, M., 2005. Electricity market modeling trends. Energy Policy, 897–913.

Weron, R., 2014. Electricity price forecasting. A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30, 1030–1081.

Weron, R., Misiorek, A., 2008. Forecasting spot electricity prices. A comparison of parametric and semiparametric time series models. International Journal of Forecasting, 24, 744–763.

Wozabal, D., Graf, C., Hirschmann, D., 2014. The effect of intermittent renewables on the electricity price variance. OR Spectrum, 1–23. DOI 10.1007/s00291-015-0395-x.

Ziel, F., Steinert R., Huisman, R., 2015. Efficient modeling and forecasting of electricity spot prices. Energy Economics, 98–111.

Zugno, M., Pinson, P., Madsen, H., 2013. Impact of Wind Power Generation on European Cross-Border Power Flows. IEEE Transactions on Power Systems.

## 6 Appendices

Figure 10 shows German electricity prices orders for every hour separately in 2015. The visible dependency of every hour of one day to its predecessor of the previous day highlights the relevance of the panel approach (see section 2.2).

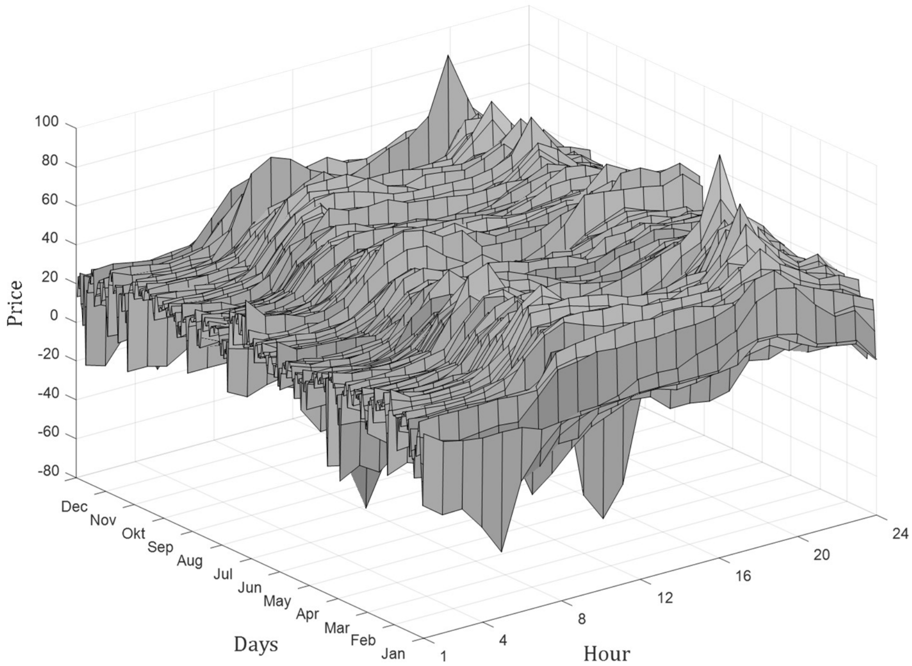


Figure 10: Surface of day-ahead prices 2015

Table 12: MWEs

| MWE | M1 | M2 | M3 | M4 | M5 | M6 |
|---|---|---|---|---|---|---|
| **1** | 0.271 | 0.279 | 0.278 | 0.265 | 0.267 | 0.269 |
| **2** | 0.140 | 0.131 | 0.130 | 0.132 | 0.131 | 0.131 |
| **3** | 0.139 | 0.134 | 0.135 | 0.132 | 0.132 | 0.131 |
| **4** | 0.089 | 0.088 | 0.088 | 0.083 | 0.085 | 0.085 |
| **5** | 0.116 | 0.113 | 0.113 | 0.114 | 0.112 | 0.111 |
| **6** | 0.101 | 0.098 | 0.099 | 0.095 | 0.098 | 0.098 |
| **7** | 0.109 | 0.104 | 0.101 | 0.105 | 0.102 | 0.101 |
| **8** | 0.095 | 0.096 | 0.097 | 0.093 | 0.095 | 0.095 |
| **9** | 0.117 | 0.109 | 0.109 | 0.110 | 0.108 | 0.106 |
| **10** | 0.125 | 0.119 | 0.120 | 0.129 | 0.126 | 0.125 |
| **11** | 0.098 | 0.093 | 0.093 | 0.094 | 0.091 | 0.090 |
| **12** | 0.115 | 0.104 | 0.104 | 0.124 | 0.116 | 0.114 |
| **13** | 0.086 | 0.084 | 0.085 | 0.083 | 0.083 | 0.083 |
| **14** | 0.187 | 0.176 | 0.176 | 0.192 | 0.180 | 0.178 |
| **15** | 0.113 | 0.108 | 0.107 | 0.105 | 0.105 | 0.104 |
| **16** | 0.162 | 0.156 | 0.156 | 0.160 | 0.155 | 0.154 |
| **17** | 0.138 | 0.128 | 0.127 | 0.140 | 0.132 | 0.132 |
| **18** | 0.168 | 0.163 | 0.162 | 0.170 | 0.169 | 0.168 |
| **19** | 0.242 | 0.237 | 0.232 | 0.245 | 0.238 | 0.239 |
| **20** | 0.217 | 0.202 | 0.201 | 0.213 | 0.207 | 0.206 |
| **21** | 0.176 | 0.175 | 0.176 | 0.179 | 0.174 | 0.174 |
| **22** | 0.193 | 0.178 | 0.174 | 0.187 | 0.176 | 0.175 |
| **23** | 0.183 | 0.178 | 0.178 | 0.184 | 0.179 | 0.18 |
| **24** | 0.149 | 0.139 | 0.137 | 0.144 | 0.138 | 0.138 |
| **25** | 0.167 | 0.156 | 0.156 | 0.169 | 0.161 | 0.161 |
| **26** | 0.139 | 0.136 | 0.135 | 0.137 | 0.134 | 0.135 |
| **27** | 0.129 | 0.125 | 0.125 | 0.135 | 0.131 | 0.13 |
| **28** | 0.125 | 0.120 | 0.119 | 0.125 | 0.120 | 0.121 |
| **29** | 0.101 | 0.100 | 0.100 | 0.104 | 0.102 | 0.102 |
| **30** | 0.139 | 0.134 | 0.133 | 0.137 | 0.131 | 0.131 |
| **31** | 0.163 | 0.160 | 0.160 | 0.161 | 0.159 | 0.16 |
| **32** | 0.158 | 0.144 | 0.143 | 0.156 | 0.144 | 0.144 |
| **33** | 0.130 | 0.124 | 0.124 | 0.124 | 0.119 | 0.121 |
| **34** | 0.140 | 0.127 | 0.128 | 0.139 | 0.130 | 0.129 |
| **35** | 0.138 | 0.131 | 0.132 | 0.133 | 0.126 | 0.127 |
| **36** | 0.143 | 0.127 | 0.126 | 0.145 | 0.130 | 0.129 |
| **37** | 0.117 | 0.113 | 0.113 | 0.113 | 0.111 | 0.111 |
| **38** | 0.161 | 0.15 | 0.151 | 0.166 | 0.155 | 0.154 |
| **39** | 0.107 | 0.095 | 0.095 | 0.115 | 0.102 | 0.103 |
| **40** | 0.167 | 0.154 | 0.152 | 0.166 | 0.156 | 0.152 |
| **41** | 0.114 | 0.101 | 0.102 | 0.116 | 0.105 | 0.105 |
| **42** | 0.136 | 0.125 | 0.124 | 0.140 | 0.132 | 0.130 |
| **43** | 0.117 | 0.108 | 0.109 | 0.120 | 0.110 | 0.111 |
| **44** | 0.124 | 0.121 | 0.119 | 0.129 | 0.123 | 0.121 |
| **45** | 0.106 | 0.099 | 0.099 | 0.106 | 0.100 | 0.101 |
| **46** | 0.140 | 0.132 | 0.128 | 0.143 | 0.135 | 0.134 |
| **47** | 0.136 | 0.132 | 0.134 | 0.136 | 0.132 | 0.133 |
| **48** | 0.101 | 0.103 | 0.100 | 0.099 | 0.098 | 0.097 |
| **49** | 0.111 | 0.109 | 0.111 | 0.114 | 0.115 | 0.116 |
| **50** | 0.119 | 0.119 | 0.118 | 0.116 | 0.11 | 0.11 |
| **51** | 0.106 | 0.106 | 0.105 | 0.113 | 0.110 | 0.111 |
| **52** | 0.321 | 0.305 | 0.307 | 0.313 | 0.302 | 0.301 |
| **53** | 0.116 | 0.119 | 0.119 | 0.119 | 0.120 | 0.121 |
| **Mean** | **0.142** | **0.135** | **0.135** | **0.141** | **0.136** | **0.136** |