Forecasting the Distribution of Hourly Electricity Spot Prices

- Accounting for Cross Correlation Patterns and Non-Normality of Price Distributions

Arne Vogler
Co-Authors: Christoph Weber, Christian Pape and Oliver Woll

Berlin, 10\textsuperscript{th} February 2017
<table>
<thead>
<tr>
<th>Agenda</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecasting the Distribution of Hourly Electricity Spot Prices</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Forecasting Approach</td>
<td>2</td>
</tr>
<tr>
<td>Evaluation of Forecast Quality</td>
<td>3</td>
</tr>
<tr>
<td>Application and Results</td>
<td>4</td>
</tr>
<tr>
<td>Conclusion</td>
<td>5</td>
</tr>
</tbody>
</table>
Weron (2014) maintains that, despite being well established in other fields of time series analysis, distribution forecasting has received little attention in electricity price forecasting.

Yet, increased production of variable RES causes higher uncertainty.

Thus, the usage of point forecasts only reduces the quality of decision making, due to the reduced amount of information provided.

Forecasting the distribution of hourly prices is more appropriate for
- the valuation of assets’ flexibilities and optionality,
- short-term decision making such as dispatch,
- and providing further information about forecast quality.
The present econometric-stochastic model combines several established approaches to adequately capture distribution characteristics.

**Panel Data**
- We model the prices of individual hours separately.

**Multiple Regression Analysis**
- We use a linear regression model to account for the deterministic components of prices and to derive the residuals.

**Mapping to Normal Distribution**
- We map the empirical cumulative distribution function of the residuals to a standard normal cumulative distribution to account for non-normality of the price distribution.
Forecasting Approach

- **Factor Model**
  - We apply a factor model to the transformed residual time series to identify common factors and to thereby account for cross correlation between hours.

- **ARMA-GARCH Class**
  - We model the time series of the identified factors using ARMA-GARCH specifications to account for autocorrelation and time-varying volatility.

- **Monte Carlo Simulation**
  - We reverse the estimation procedure using Monte Carlo simulations to derive prediction samples.

→ We essentially characterize the distribution of $x_{t,h}$ (the price of hour $h$ at day $t$) as the empirical cumulative distribution function of a Monte
Given a sample \( \{y_t, I_t\}_{t=1}^T \), we seek to test whether \( y_t | I_t \) has a specific parametric form.

Thus, we wish to test the following null hypothesis

\[
H_0: \Pr(Y_t \leq y | I_t, \theta_0) = F_t(y | I_t, \theta_0)
\]

That is, we seek to assess calibration.

The evaluation of distribution forecasts rests on the probability integral transform (PIT), also known as Rosenblatt transformation (1952).

- Under the null hypothesis \( F_t(y_t | I_t, \theta_0) \) follows a uniform distribution on \([0,1]\).
- Additionally, the PIT values from a k-step-ahead forecast should be at most (k-1)-dependent, depending on information set \( I_t \).
- The PIT values of the distribution forecasts, \( F_t(y_t | I_t, \hat{\theta}_T) \), over a hold-out sample can be used to assess calibration.
The Evaluation Framework (II)

Evaluation of Forecast Quality

- The graphical evaluation framework
  - The classic econometric testing framework rests on a graphical analysis of these PIT values.
  - Histogram and Sample Autocorrelation Function
  - Yet, it should be noted that (k-1) dependence hinges crucially on $I_t$ being equal to the “relevant” information set.

- Evaluation and formal tests
  - Depending on the information set, the PIT values may exhibit autocorrelation, which formal tests have to account for.
  - Thus, classic Kolmogorov-type tests that rely on i.i.d. observations cannot be applied.
  - Knüppel (2015) proposes a test that is robust to autocorrelation and for which standard critical values can be used.
An alternative evaluation framework

- The probabilistic forecasting test framework rests mainly on the evaluation of the uniformity of the PIT values (graphically and formally), sharpness and various scores measures.
- The proposed paradigm is to minimize sharpness subject to calibration, where sharpness is a characteristic of the forecast only and refers to the concentration of the distribution forecast.

→ Calibration constitutes a necessary but not sufficient condition for an ideal distribution forecast. We thus require the PIT values to be at least uniformly distributed.

→ Any dependence patterns may shed light on the characteristics of the information set underpinning our specification.
Application and Results

- We test our econometric-stochastic approach against German day-ahead prices for 2014 and 2015 separately.

- We consider 12 different specifications.
  - ARMA-GARCH Class: AR(1), AR(2) and ARMA(1,1)-GARCH(1,1)
  - Factor Model: on and off
  - Sample Size: 730 and 184

- We calculate daily out-of-sample day-ahead forecasts using a rolling window for 2014 and 2015; thus, running 8760 Monte Carlo price simulations for each year and specification.

- Based on the evaluation framework, we conclude …
  - … the AR(2) model with the factor model to work best for 2014
  - … the AR(2) model without the factor model to work best for 2015
We fail to reject the null hypothesis of calibration for 22 hours of 2015 under the preferred specification.
We fail to reject the null hypothesis of calibration for 19 hours of 2014 under the preferred specification.
### Application (IV)

#### Application and Results

- The formal calibration tests, due to Knüppel (2015), confirms the results of the preceding graphical analysis.

| Subsample | Method          | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | Sum |
|-----------|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 2014      | PCA0_AR2_730    | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 6  |
|           | PCA1_AR2_730    | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 9  |
|           | PCA0_AR2_184    | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 12 |
|           | PCA1_AR2_184    | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 19 |
| 2015      | PCA0_AR2_730    | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 22 |
|           | PCA1_AR2_730    | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 20 |
|           | PCA0_AR2_184    | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 1  | 1  | 0  | 1  | 0  | 20 |
|           | PCA1_AR2_184    | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 1 | 1  | 1  | 0  | 1  | 1  | 1  | 1  | 1  | 15 |

- The present econometric-stochastic approach delivers calibrated distribution forecasts.
Application and Results

- Yet, the analysis of the sample autocorrelation function uncovers violation of the at most \((k-1)\) dependence criterion for 2015.
The econometric-stochastic approach is able to capture the main characteristics of daily hourly prices in Germany and delivers calibrated distribution forecasts.

A few comments on model particularities are warranted

- Factor models adequately address cross correlations and ensure smooth price paths

- Time-varying volatility seems to be less important for price processes of individual hours, as GARCH specifications do not improve results

- The conditional distributions are correctly specified with respect to the considered information set; yet, dynamic misspecification seems to be present.
Thank you for your attention!

Arne Vogler
House of Energy Markets & Finance
Universität Duisburg-Essen
Weststadttürme | Berliner Platz 6-8
45127 Essen
(i) Determine the main deterministic drivers and the residuals

- Regression model to account for deterministic factors

\[ x_{t,h} = \beta_0 + \beta_1 (L_{t,h} - S_{t,h}) + \beta_2 (W_{t,h}) + \beta_3 (C_{coal,t}) + \beta_4 (C_{Gas,t}) + \varepsilon_{t,h} \]

- \( \beta \): Regression coefficients
- \( L_{t,h} \): Load
- \( S_{t,h} \): Solar
- \( W_{t,h} \): Wind
- \( C_{coal,t} \) and \( C_{Gas,t} \): typical variable costs of power plants incl. emission costs
- \( \varepsilon_{t,h} \): Residuals
(ii) Map the empirical CDF of residuals onto a normal distribution

\[ T_h: \varepsilon_{t,h} \mapsto \Phi^{-1}\left( C_h(\varepsilon_{t,h}) \right) \]

- \( C_h \): Empirical CDF of residuals in hour \( h \)
- \( \Phi \): CDF of the standard normal distribution
- Graphical representation corresponds to Q-Q-plot (Quantile Mapping)

(iii) Factor Model

- Normal Residuals

\[ u_{t,h} = T_h(\varepsilon_{t,h}) \]

- The common factors are constructed using a principal component analysis on the correlation matrix of the transformed factors.
(iv) Factor time series are modelled with ARMA-GARCH specifications

- **ARMA (1,1):**
  \[ f_{t,i} = \alpha_{1,i} f_{t-1,i} + \alpha_{2,i} w_{t-1,i} + w_{t,i} \]
  - \( \alpha \): Coefficients for ARMA part
  - \( f_{t,i} \): Factors i from different time steps t
  - \( w_{t,i} \): Error term from different time steps and \( w_{t,i} \sim N(\mu, \sigma) \)

- **GARCH (1,1):**
  \[ \sigma_{t,i}^2 = \gamma_{0,i} + \gamma_{1,i} \sigma_{t-1,i}^2 + \gamma_{2,i} w_{t-1,i}^2 \]
  - \( \gamma \): Coefficients for GARCH part \( (\gamma_0 = \text{constant term}) \)
  - \( w_{t-1,i}^2 \): Error term from the previous time step
  - \( \sigma_{t-1,i}^2 \): Volatility from the previous time step

(v) Maximum Likelihood Estimation of parameters using a rolling window

(vi) Price distributions are simulated by performing the steps (i) to (v) in reverse order
Q-Q-Plot

- Quantile Mapping

![Graph showing Q-Q plot with empirical quantiles on the x-axis and standard normal quantiles on the y-axis. The graph shows a curve that deviates slightly from the straight line, indicating some deviation from normality.](image-url)
Orthogonal Factor Model

- The Orthogonal Factor Model (Johnson and Wichern (2002))

\[ X = \mu + LF + \varepsilon \]

implies a specific covariance structure

\[ \Sigma = LL' + \Psi, \quad \Psi = \text{Cov}(\varepsilon) \]

which can be used to solve for factor loadings \( L \) and common factors \( F \) by spectral decomposition.

\[ \Sigma = \left[ \sqrt{\lambda_1} e_1, \ldots, \sqrt{\lambda_p} e_p \right] \left[ \sqrt{\lambda_1} e_1, \ldots, \sqrt{\lambda_p} e_p \right]' \]

\[ L \]