# Forecasting day-ahead electricity prices: utilizing hourly prices*

Eran Raviv[1,2,*], Kees E. Bouwman[1], and Dick van Dijk[1,2,3]

[1]*Econometric Institute, Erasmus University Rotterdam*
[2]*Tinbergen Institute*
[3]*Erasmus Research Institute of Management*

First draft - January 17, 2012
This version - May 6, 2013

## Abstract

The daily average price of electricity represents the price of electricity to be delivered over the full next day and serves as a key reference price in the electricity market. It is an aggregate that equals the average of hourly prices for delivery during each of the 24 individual hours. This paper demonstrates that the disaggregated hourly prices contain useful predictive information for the daily average price. Multivariate models for the full panel of hourly prices significantly outperform univariate models of the daily average price, with reductions in Root Mean Squared Error of up to 16%. Substantial care is required in order to achieve these forecast improvements. Rich multivariate models are needed to exploit the relations between different hourly prices, but the risk of overfitting must be mitigated by using dimension reduction techniques, shrinkage and forecast combinations.

*Keywords:* Electricity market; Forecasting; Hourly prices; Dimension reduction; Shrinkage; Forecast combinations.
*JEL-code:* C53; C32; Q47

# 1   Introduction

Over the last decades, electricity markets in many developed countries have experienced an ongoing liberalization, such that nowadays prices generally are determined by the forces of supply and demand. Electricity shows unique price behavior, largely due to distinct characteristics compared to other traded commodities. In particular, electricity is a *flow* commodity in the sense that is virtually non-storable. Shocks to demand or supply therefore cannot be smoothed using pre-stored inventory unlike for other, *stock* commodities. As a result, electricity prices exhibit high volatility, much higher than other energy products, which leads to significant price risk for market participants. Other unique characteristics include multiple levels of seasonality, price spikes, mean reversion, and the possibility of negative prices (Knittel and Roberts, 2005; Longstaff and Wang, 2004). At the same time, contracts traded on electricity such as forwards and futures differ from those for other commodities. For example, electricity forward contracts concern delivery over an extended period rather than at a single point in time.

The unique features of electricity prices (and related financial contracts) have spurred the demand for econometric models that can adequately describe their dynamics, see Benth and Koeke-bakker (2008) and Möst and Keles (2010) for surveys. From a practical point of view, models of electricity prices are crucially important for forecasting, derivatives pricing and risk management. Several papers examine dynamic models of electricity prices for forecasting purposes, see Zareipour (2012) for a recent review. Weron and Misiorek (2008) forecast daily electricity prices using a variety of linear and non-linear time-series models, including basic autoregressive models, jump-diffusion models and regime-switching models, see also Weron (2006). Karakatsani and Bunn (2008) add fundamental variables such as fuel prices and level of demand, although their focus is mostly on accurately modeling the volatility of electricity prices. In a similar vein, Huurman *et al.* (2012) examine the added value of information from temperature levels for predicting daily electricity prices. Christensen *et al.* (2012) focus on predicting the occasional extreme spikes in electricity prices using an autoregressive conditional hazard model.

The majority of electricity price models and related studies on forecasting focus on the *daily average price*. This is not surprising as the daily average price plays a central role in the electricity market. It acts as a proxy for the spot price of electricity and as a reference price for forward and futures contracts as well as many other derivatives contracts. In various electricity markets, including the Nord Pool Spot system that we consider in this paper, the daily average price is established in the so-called day-ahead market. This concerns an auction market where participants

trade electricity for delivery during the individual hours of the next day. Quotes for all hours are submitted *simultaneously* and market prices are determined by the intersection point of the aggregate demand and supply curves. The daily average price is equal to the average of the 24 individual hourly prices.

This paper addresses the question whether the electricity prices for individual hours contain useful predictive information for the daily average price. This question is important for several reasons. Forecasts for the daily average price are highly relevant for market participants, given its crucial importance for trading and derivatives pricing. Improved forecasts can help market participants to develop more efficient trading and bidding strategies to increase profits and control risk. The hourly prices provide a rich source of information that can potentially help to improve forecasts. The question has also important implications for our understanding of electricity price dynamics and the way these are best modeled. The standard approach, where the time series of the daily average price is modeled directly, ignores potentially relevant information in intraday hourly prices.

Hourly prices have received quite limited attention in the literature, in particular when it comes to forecasting. Cuaresma *et al.* (2004) and Kristiansen (2012) use different autoregressive time-series models for forecasting individual hourly prices. Boogert and Dupont (2008) employ a more structural approach based on a supply-demand framework, and also consider probability forecasts for price spikes.

An important consideration that is often overlooked in modeling (and forecasting) intraday prices is that typically in practice, the prices of all the hours of the next day are determined *simultaneously*. Hence, it is not appropriate to model the hourly prices as a single time series. The price for the hour between 20:00-21:00h, for example, cannot be exploited to predict the price for the subsequent hour 21:00-22:00h, as these two prices are determined simultaneously. Instead, hourly prices should be modeled as a panel of 24 daily time-series of hourly prices, as argued by Huisman *et al.* (2007) and Härdle and Trück (2010), among others.

We conduct an empirical forecasting exercise to address the question whether hourly electricity prices contain predictive information for the daily average price. We use price data from the Nord Pool Spot exchange, for a sample period covering almost two decades from 1992 to 2010. We compare forecasts of the daily average price based on univariate time-series models with forecasts derived from multivariate time-series models for the full panel of prices for the 24 individual hours.

Effectively exploiting information in hourly prices is challenging as modeling all individual hours separately quickly increases model complexity and the number of parameters. For example, model-

ing the panel of all 24 hours by a standard first-order Vector Autoregressive (VAR(1)) model with intercepts will produce point forecasts depending on 600 unknown parameters that have to be estimated. Such a large number of parameters effectively dissipates our degrees of freedom, resulting in large estimation uncertainty, potential overfitting of in-sample data and increased variability of out-of-sample forecast errors.

We address this 'curse-of-dimensionality' problem in three different ways. First, we limit model complexity by applying dimension reduction techniques such as Reduced Rank Regression and Factor Models, leading to a substantial decline in the number of parameters. Second, regularization, or shrinkage, is used to obtain parameter estimates that are less prone to overfitting. In particular, we consider a Bayesian VAR model with a prior distribution on the parameters that shrinks the model towards a random walk process for each individual hour. Third, forecasts of different models with different levels of complexity are combined. It is well-established that combining forecasts of different models may well perform better than the best individual forecast (Timmermann, 2006). Overfitting will have a different effect on the individual models and hence combining their forecasts might potentially average out these effects.

Our empirical results show that hourly prices contain substantial predictive information for the daily average price. Multivariate models achieve forecast improvements up to 15% in out-of-sample Root Mean Squared Error (RMSE) over a flexible univariate benchmark model. An effective use of this predictive information however requires rich models that account for the complex relations between prices of different hours. Moreover, dimension reduction, shrinkage and forecast combinations lead to further improvement in forecasting performance.

The rest of the paper is organized as follows. We describe the data set in Section 2. In Section 3 we introduces the forecasting models, with a detailed exposition of the various techniques that we employ to model the full panel of hourly prices while avoiding overfitting and related issues. In Section 4 we present the empirical results, discuss their significance and assess their robustness. We conclude in Section 5.

## 2 Data Analysis

We use data stemming from the Nordic power exchange, Nord Pool Spot, owned by the Nordic and Baltic transmission system operators and operating the leading power markets in Europe. About 370 companies from 20 countries trade on the Nord Pool Spot's markets, with participants including both producers and large consumers. We consider electricity prices as determined in the Elspot

market. This is the largest of Nord Pool Spot's markets, with a trading volume of approximately 330 terawatt hours in 2012, representing 77% of Nordic power consumption. Specifically, it comprises Norway, Sweden, Finland, Denmark and (since 2010) Estonia and (since 2012) Lithuania.

Elspot is an auction market for delivery the following day. The daily auction procedure works as follows. Participants submit bids and offers for each individual hour of the next day, through the Nord Pool Spot web-based trading system. Orders can be placed until 12:00h Central European Time (CET). Buy and sell orders are then aggregated into demand and supply curves for each delivery hour. The system price for each hour is determined by the intersection of these curves, also taking into account the transmission capacity of the power system. Prices are quoted for megawatt hour (MWh). Our data set consists of the twenty-four hourly prices for each day for the period from May 4, 1992 up to March 4, 2010, covering 6519 observation days. All prices are converted to Norwegian krone (NOK).
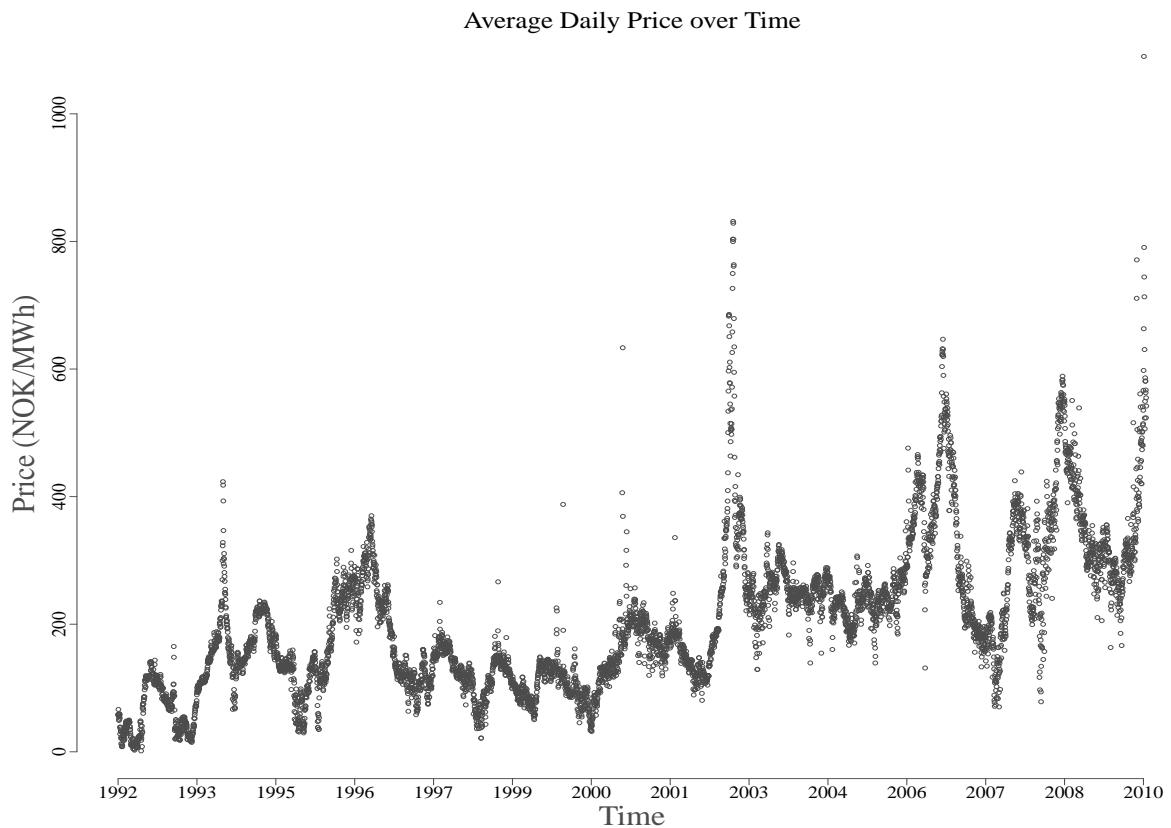


**Figure 1:** Daily Average Price

Figure 1 shows the time series of the daily average price, which is the main object of interest in our empirical forecasting exercise. The daily average price shows substantial fluctuations, between minimum values close to zero up to occasional spikes at around 800 NOK/MWh.

| Hour | Mean | St.Dev. | Skewness | Kurtosis | $\widehat{\rho}(1)$ | $\widehat{\rho}(2)$ | $\widehat{\rho}(7)$ |
|---|---|---|---|---|---|---|---|
| 1 | 194.23 | 111.40 | 1.07 | 4.89 | 0.99 | 0.99 | 0.96 |
| 2 | 188.73 | 109.34 | 1.07 | 4.87 | 0.99 | 0.99 | 0.95 |
| 3 | 185.03 | 108.48 | 1.10 | 5.02 | 0.98 | 0.98 | 0.93 |
| 4 | 182.29 | 107.09 | 1.08 | 4.95 | 0.98 | 0.98 | 0.94 |
| 5 | 183.12 | 108.54 | 1.12 | 5.09 | 0.98 | 0.98 | 0.94 |
| 6 | 189.36 | 111.52 | 1.11 | 5.07 | 0.98 | 0.98 | 0.94 |
| 7 | 199.10 | 114.84 | 1.08 | 4.91 | 0.97 | 0.97 | 0.95 |
| 8 | 211.83 | 123.76 | 1.47 | 8.71 | 0.93 | 0.93 | 0.90 |
| 9 | 221.98 | 140.12 | 3.09 | 33.68 | 0.82 | 0.82 | 0.77 |
| 10 | 221.74 | 132.29 | 2.31 | 22.44 | 0.88 | 0.88 | 0.84 |
| 11 | 221.33 | 126.88 | 1.53 | 9.75 | 0.94 | 0.94 | 0.89 |
| 12 | 219.47 | 123.52 | 1.26 | 6.43 | 0.96 | 0.96 | 0.93 |
| 13 | 215.94 | 120.78 | 1.14 | 5.39 | 0.98 | 0.98 | 0.95 |
| 14 | 213.46 | 119.08 | 1.11 | 5.11 | 0.98 | 0.98 | 0.95 |
| 15 | 211.57 | 118.16 | 1.11 | 5.08 | 0.98 | 0.98 | 0.95 |
| 16 | 210.71 | 118.42 | 1.12 | 5.01 | 0.97 | 0.97 | 0.95 |
| 17 | 212.37 | 122.95 | 1.32 | 6.42 | 0.95 | 0.95 | 0.92 |
| 18 | 217.25 | 131.10 | 1.72 | 10.18 | 0.92 | 0.92 | 0.87 |
| 19 | 216.52 | 126.81 | 1.40 | 7.21 | 0.96 | 0.96 | 0.91 |
| 20 | 212.72 | 120.84 | 1.15 | 5.32 | 0.98 | 0.98 | 0.95 |
| 21 | 209.26 | 117.25 | 1.08 | 4.84 | 0.99 | 0.99 | 0.96 |
| 22 | 207.96 | 115.85 | 1.07 | 4.85 | 0.99 | 0.99 | 0.96 |
| 23 | 204.89 | 114.01 | 1.04 | 4.66 | 0.99 | 0.99 | 0.96 |
| 24 | 197.20 | 110.49 | 1.05 | 4.72 | 0.99 | 0.99 | 0.96 |
| | | | | | | | |
| Daily Average | 206.16 | 116.63 | 1.13 | 5.21 | 0.98 | 0.96 | 0.94 |

**Table 1:** The table presents summary statistics for Elspot's day-ahead prices for individual hours of the day and the daily average price for the period from May 4, 1992 up to March 4, 2010 (6519 days). Hour 1 runs from 00:00-00:59h Central European Time (CET), etc. Prices are quoted in Norwegian krone (NOK). During the sample period, one euro was approximately 8.5 NOK. $\widehat{\rho}(k)$ is the $k$-th order sample autocorrelation.

Table 1 presents descriptive statistics of the hourly and daily average price series. Electricity prices are on average higher during the day than during the night, peaking in the morning (08:00 - 12:00h CET) and early evening (17:00-19:00h CET). These peak hours also have relatively high skewness and kurtosis and are less persistent, reflecting that these hours experience more extreme prices and price spikes. This also explains the relatively high volatility during peak hours. Note that electricity prices have substantial volatility throughout the day: for all hours the standard deviation is about half the mean price.

Figures 2 and 3 visually depicts some of the salient features of the electricity prices for individual hours. The top and bottom panels of Figure 2 display the weekly and annual seasonality in hourly prices. From the top panel we observe that weekdays have higher prices on average than Saturdays and Sundays, but also that the two weekend days have a slightly different intraday price pattern. It is also interesting to note that on Friday afternoon prices already start to decline towards the level of Saturday and Sunday, reflecting the early close of offices and factories on that day. For the annual seasonality we observe that, as expected, winter months experience higher prices than summer months. Average prices in August are higher already than those in June and July, presumably due to the use of energy-intensive airconditioning systems.

The cross-correlation structure shown in Figure 3 reveals that prices during the afternoon and nightly hours vary closely together while prices in the early morning and early evening hours show less strong comovement. For example, prices for hour number 9 (between 08:00h and 09:00h CET) have relatively weak correlation with the rest of the day. Overall, however, the correlation is high, between 0.85 and 0.99.

The very substantial cross-correlations between the individual hourly prices indicate that the fluctuations in these prices might be effectively summarized by a limited number of common factors. We examine this by performing a Principal Component Analysis (PCA) on the full panel of 24 hourly prices based on its correlation matrix. Figure 4 presents the time series of the first two principal components along with their corresponding loadings for the individual hourly prices. Not surprisingly, the first principal component can be interpreted as a level factor that captures the fluctuations in the overall level of electricity prices common to all hours. Based on its loadings, the second principal component can be interpreted as a spread factor between prices of peak hours (08:00-12:00h CET and 17:00-19:00h CET) and off-peak hours. The first principal component captures about 96.2% of the total variance of the panel, whereas the second principal component adds another 2.2%. The remaining 22 components thus account for the remaining 1.6% of the variance in the panel. These results demonstrate that indeed the bulk of the fluctuations and
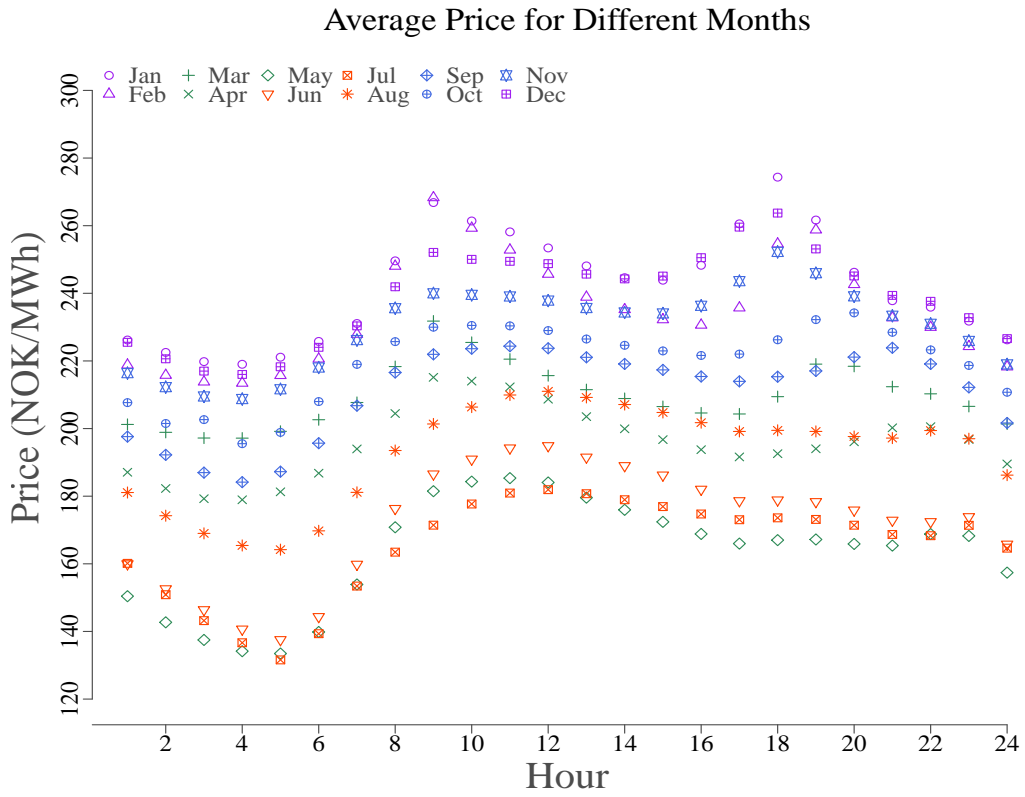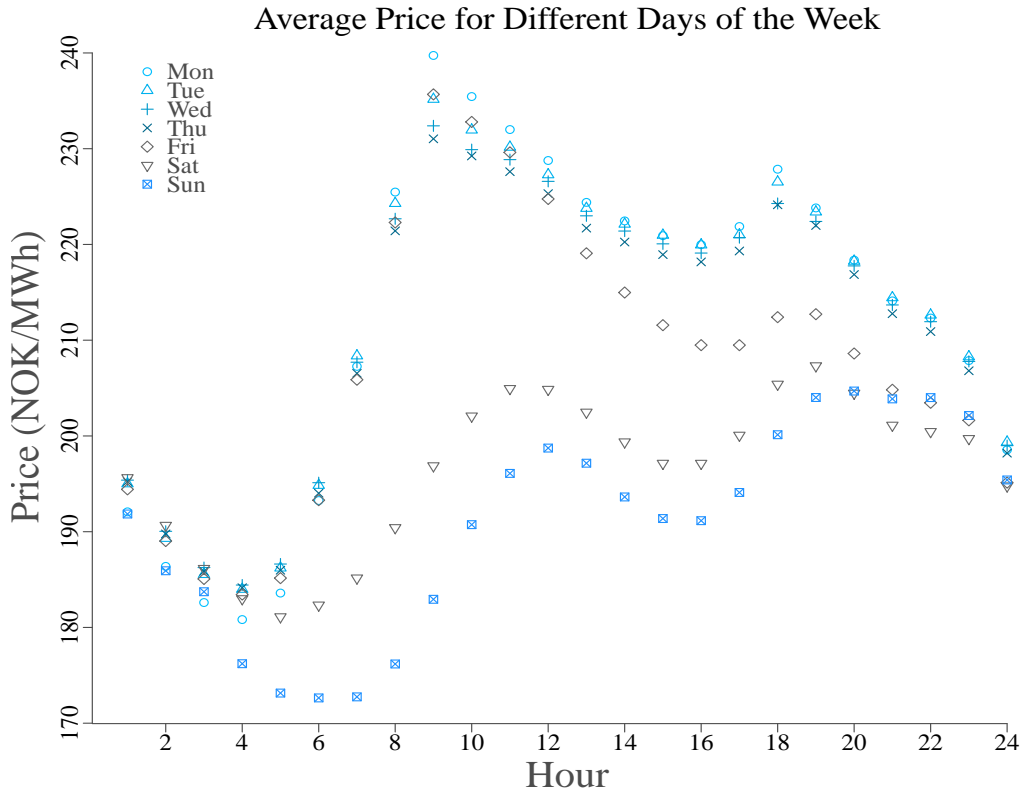
**Figure 2:** **Top**: Average of hourly prices for different days of the week. **Bottom**: Average of hourly prices for different months, with the different colors representing different seasons.
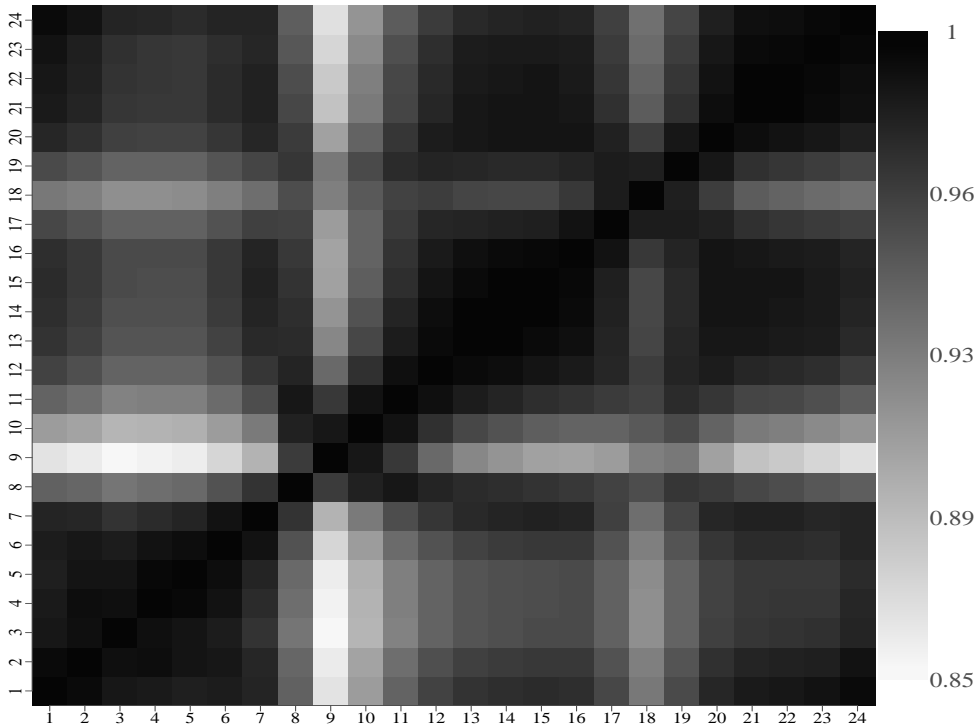
Hourly Prices Cross-Correlation Structure



**Figure 3:** Hourly Prices Cross-Correlation Structure

comovement in the panel of hourly prices can effectively be represented by means of a small number of factors. This encourages the use of dimension reduction techniques for forecasting, such as, e.g., Principal Component Regression (PCR).

# 3    Forecast Methods

In this section we present the forecast methods we consider for predicting the daily average electricity price. We compare univariate forecasting models for the daily average price itself with several multivariate models for the full panel of hourly prices. From the multivariate models, we obtain forecasts for all hourly prices, which are averaged to deliver a forecast of the daily average price.

To fix notation, let $y_{ht}$ denote the price for hour $h$ on day $t$. The daily average price is given by

$$\bar{y}_t = \frac{1}{H} \sum_{h=1}^{H} y_{ht}, \qquad \text{with } H = 24. \tag{1}$$

We start with the description of univariate models for the daily average price $\bar{y}_t$ and then continue with multivariate models for the panel of hourly prices $y_{ht}$. We end this section with outlining the methods we use to combine forecasts from different models.
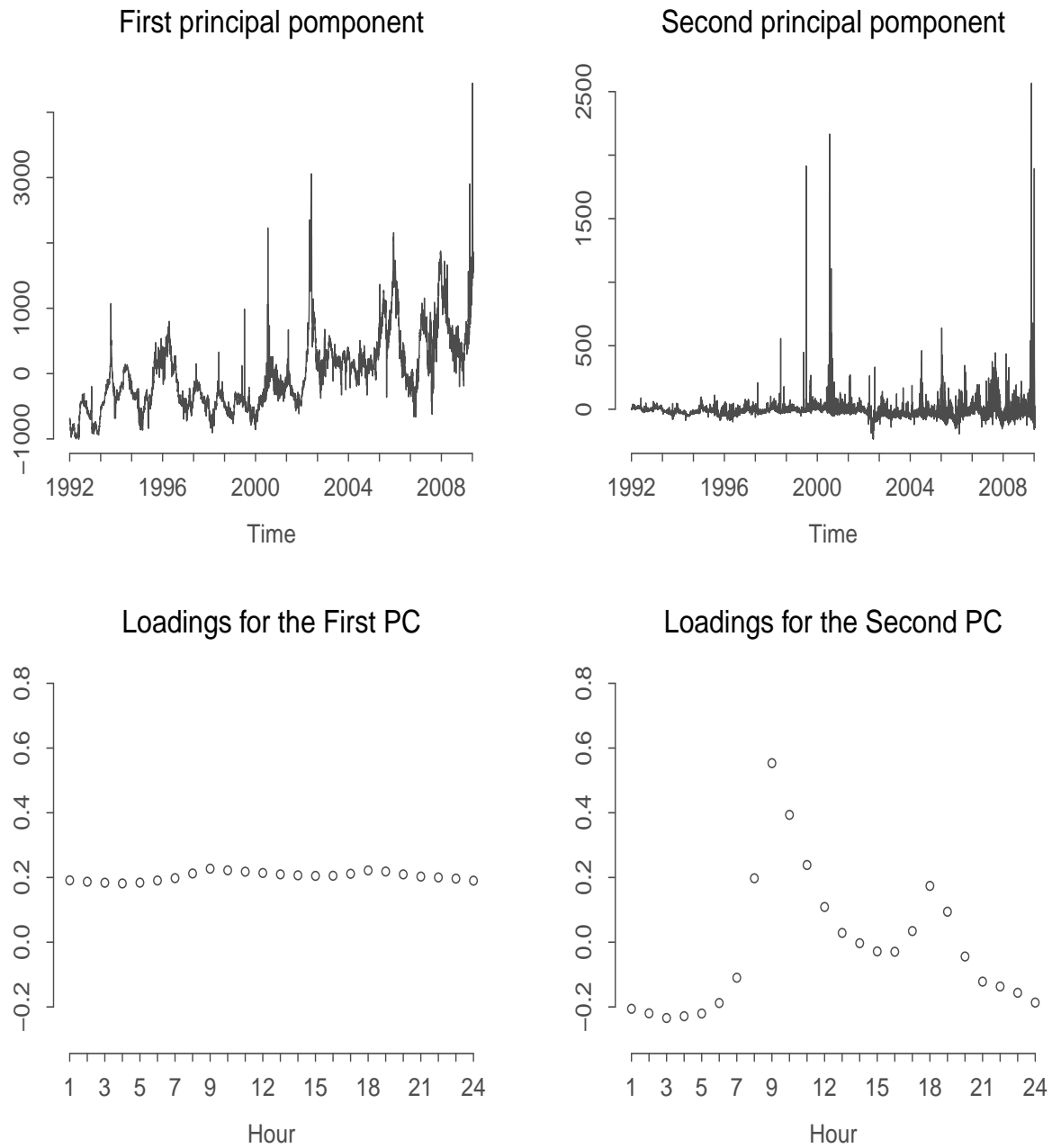
8

**Figure 4:** The figure shows results of a PCA on the panel of 24 hourly prices based on its correlation matrix. The graphs in the top panels show time series of the first and second principal components. The graphs in the bottom panels show the corresponding loadings on the individual hourly prices.

## 3.1 Univariate Models

To address the question whether the prices for individual hours contain useful predictive information for the daily average price, we compare forecasts from multivariate models for the panel of individual hours against forecasts of univariate models for the daily average price. Obviously, a multivariate framework is potentially much more flexible and comprehensive than a univariate forecasting approach. For this reason it is important to compare the multivariate forecasts against reasonable univariate benchmarks, that is, forecasting models that should be able to capture the key features of electricity prices outlined in the previous section, namely mean reversion, multiple seasonality patterns and strong autocorrelation. At the same time, we intend to focus on forecast methods that are feasible and sensible choices for practitioners. We consider two models that potentially satisfy both criteria, namely a dynamic AR model and the Heterogeneous Autoregressive model (HAR).

### 3.1.1 Dynamic AR model

An AR model of order $p$ for the daily average price is defined as:

$$\bar{y}_t = \sum_{j=1}^{p} \phi_j \bar{y}_{t-j} + \sum_{k=1}^{K} \psi_k d_{kt} + \varepsilon_t, \tag{2}$$

where $p$ is the number of lags included in the model, and $d_{kt}$ are dummies for Saturdays, Sundays and for each month of the year. We consider the AR model with a maximum of $p = 14$ lags. It is dynamic in the sense that $p$ is chosen at every point in time according to the Akaike Information Criterion (AIC). That is, at each point in time we choose the number of lags that minimizes $AIC(r) = \log \widehat{\sigma}^2(r) + 2r/N$, where $r$ is the number of regressors in the model, $\widehat{\sigma}^2(r)$ is the estimated residual variance value, and $N$ is the length of the estimation window.[1]

### 3.1.2 Heterogeneous Autoregressive model (HAR)

The Heterogeneous Autoregressive model (HAR) is an AR-type model recently proposed by Corsi (2009) to forecast realized volatility measures. The model is explicitly designed to capture long memory behavior, that is, large and slowly declining autocorrelations, which is one of the key features also of electricity prices, see Table 1. The HAR model essentially is a high order AR($p$) model as given in (2) but with restrictions on the autoregressive coefficients $\phi_j$, $j = 1, \ldots, p$. Hence, it provides a more parsimonious framework, which can be a strong advantage from a forecasting

---

[1]We also consider the more conservative Bayesian Information Criterion (BIC). Results based on the AIC proved more accurate so we make those our benchmark for comparison.

perspective. The model specification is given by

$$\bar{y}_t = \phi_1 \bar{y}_{t-1} + \phi_2 \bar{y}_{t-1,w} + \phi_3 \bar{y}_{t-1,m} + \sum_{k=1}^{K} \psi_k d_{kt} + \varepsilon_t, \tag{3}$$

where $\bar{y}_{t-1,x} = (\bar{y}_{t-1} + \cdots + \bar{y}_{t-x})/x$ is the average price during the past $x$ days, $x = \{w, m\}$. Following Corsi (2009), we use $w = 7$ and $m = 30$, corresponding with the average price over the past week and month, respectively.

## 3.2 Multivariate Models

Multivariate models for the full panel of individual hourly electricity prices obviously offer a great deal of flexibility. Potentially they allow us to exploit the rich source of information in the intraday price behavior to obtain superior forecasts of the daily average price. At the same time, an unconstrained multivariate model might be ill-behaved when it comes to forecasting. The extra flexibility may create in-sample overfitting at the expense of forecast accuracy. In particular, unconstrained multivariate models typically require the estimation of a large number of parameters. The associated estimation uncertainty will likely adversely affect the accuracy of out-of-sample forecasts. These problems are especially acute in our context as the daily average price is an aggregation of no less than 24 hourly series. For this reason, in our choice of multivariate forecasting models we explicitly focus on methods that limit the model complexity and keep the number of unknown parameters within reasonable limits. This is achieved by applying either shrinkage methods or dimension reduction techniques (or both), as discussed in the following sections.

### 3.2.1 VAR Models

Our starting point for the multivariate forecasting models is a vector autoregressive (VAR) model for the panel of hourly prices. This is a natural choice for electricity price forecasting, given their strong persistence. Define the $(H \times 1)$ vector of hourly prices as $Y_t = (y_{1t}, \ldots, y_{Ht})$ with $H = 24$, $D_t = (d_{1t}, \ldots, d_{Kt})$ a $(K \times 1)$ vector with $d_{kt}$ representing dummy variables for Saturdays, Sundays and months of the year, and finally $X_t = (Y_{t-1}, \ldots, Y_{t-p}, D_t)$. We can now write the $VAR$ model of order $p$ as

$$Y_t = \Phi' \boldsymbol{X}_t + e_t, \tag{4}$$

where the $((Hp+K) \times H)$ matrix $\Phi$ contains the autoregressive coefficients as well as the coefficients for the dummy variables, and the errors $e_t$ are assumed to be serially uncorrelated and normally

distributed with mean zero and covariance matrix $\Sigma$. It is convenient to rewrite the model in a more compact form:

$$\boldsymbol{Y} = \boldsymbol{X}\Phi + \boldsymbol{E} \tag{5}$$

where $\boldsymbol{Y} = (Y_1', \ldots, Y_R')$ is an $(R \times H)$ matrix[2], $\boldsymbol{X} = (X_1, \ldots, X_R)'$ is the $(R \times (Hp + K))$ matrix of explanatory variables, and $\boldsymbol{E} = (e_1', \ldots, e_R')$ is the $(R \times H)$ error matrix.

With $H = 24$ hours per day and $K = 14$ dummy variables, the number of coefficients in the VAR model in (4) rapidly increases to unreasonable numbers when the autoregressive order $p$ increases. A first restriction that we impose throughout is to include only the first, second and the seventh lag of the hourly prices. This choice is motivated by the strong persistence observed in electricity prices (Knittel and Roberts, 2005), as well as the weekly seasonality pattern (see the top panel of Figure 2). The same lag structure was also considered in Weron and Misiorek (2008). With a slight abuse of notation, in the remainder $p$ is used to denote the number of included lags, that is, $p = 3$, instead of the maximum lag in the VAR model.

We consider three specific VAR models. First, we consider an unrestricted VAR (UVAR), which imposes no restrictions on the coefficients in (4) at all. The UVAR model has a large number of parameters: with $p = 3$ lags and $K = 14$ exogenous dummy variables, we have $H(Hp + K) = 2064$ coefficients to be estimated. Consequently, the model is prone to overfitting and will necessarily involve a large amount of estimation uncertainty. The second model strongly limits the number of unknown parameters by restricting all coefficients of cross lags[3] to zero. The resulting model, labeled Diagonal VAR (DVAR), contains $H(p + K) = 408$ unknown parameters. Note that the DVAR model essentially boils down to a collection of 24 univariate autoregressive models for the individual hourly prices. We will see that despite the flexibility of this model where every hour is modeled separately, it does not significantly outperform the dynamic univariate AR benchmark model. The third model is a Bayesian VAR (BVAR), which uses shrinkage to limit the estimation uncertainty in the UVAR. The specific details of this approach are discussed in the following subsection.

### 3.2.2 Bayesian VAR

We use the BVAR as a practical shrinkage (regularization) device to help mitigate overfitting, not as a tool for conducting formal Bayesian inference, though such inference is possible. Define $\boldsymbol{\alpha} = vec(\Phi)$, and $\boldsymbol{y} = vec(\boldsymbol{Y})$, where $vec(\cdot)$ is the usual vec operator. We can now rewrite the model

---

[2]Here and throughout we assume that pre-sample values $Y_0, Y_{-1}, \ldots, Y_{1-p}$ are available such that the sample covers $R$ observations.

[3]That is, coefficients of $y_{i,t-j}$ for $i = 1, 2, \ldots, 24$, $i \neq h$, $j = 1, 2$ and 7, in the equation for $y_{ht}$.

as:

$$\boldsymbol{y} = (\boldsymbol{I}_H \otimes \boldsymbol{X})\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{6}$$

where $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_R)$, and $\boldsymbol{I}_{(.)}$ is an identity matrix.

The BVAR approach limits the problems of overfitting and estimation uncertainty in the uncon-strained model (6) by constraining the coefficients. Unlike the DVAR model, however, the coefficient restrictions are not 'strict', but more subtle in the form of shrinkage or regularization. Specifically, we shrink the coefficient estimates by combining their unconstrained least squares estimates with a certain prior distribution. We follow the conventional approach and choose a Minnesota prior distribution with mean and variance such that the estimates are shrunk towards a random walk specification for the individual hourly prices. In general, the Minnesota prior assumes that $\boldsymbol{\alpha}$ is normally distributed with prior mean $\boldsymbol{\alpha}^{prior}$ and covariance matrix $\boldsymbol{V}^{prior}$. For the elements of $\boldsymbol{\alpha}^{prior}$, we use a value of one for the coefficients of the first order 'own' lag $y_{h,t-1}$ in the equation for $y_{ht}$. Coefficients for own lags of orders beyond one and all cross lags are shrunk towards zero. In terms of the original coefficient matrix $\Phi$, this boils down to

$$\Phi_{ij}^{prior} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where $\Phi_{ij}$ denotes the $(i,j)$-th element of $\Phi$. The random walk prior is the traditional and most common choice for the prior means, see Koop and Korobilis (2010), among many others.[4]

The prior covariance matrix $\boldsymbol{V}^{prior}$ determines the amount of tightness around the prior mean. The larger the prior variance, the closer the final estimate is to its unrestricted VAR estimate. Here we do not impose shrinkage on the coefficients of the exogenous variables by assigning large values to their prior variances, such that their posterior values essentially are equal to the unrestricted OLS estimates. Furthermore, we follow the convention to assume smaller variances for higher order lags, reflecting that these should have smaller overall impact in prediction. The Minnesota prior assumes the prior covariance matrix to be diagonal. Let $\boldsymbol{V}_h$ be the block associated with the coefficients in equation $h$, and let $\boldsymbol{V}_{h,ii}$ be its diagonal elements, $i = 1, \ldots, Hp + K$. We specify the prior variance

---

[4]We can also rely on the fact that energy prices are mean reverting, and use a value smaller than one for the first order 'own' lag coefficients. We do not follow this path to avoid a somewhat arbitrary choice.

of the coefficients in the equation variable $h$ as:

$$
\boldsymbol{V}_{h,ii} =
\begin{cases}
\frac{\lambda_1}{l^2} & \text{for coefficients on own lags for lag } l = 1, \ldots, p \\[2mm]
\frac{\lambda_2}{l^2} \frac{\sigma_i}{\sigma_h} & \text{for coefficients on cross lags of } y_{it} \text{ for lag } l = 1, \ldots, p \\[2mm]
\lambda_3 \sigma_h & \text{for coefficients on exogenous dummy variables}
\end{cases}
\tag{8}
$$

We estimate $\sigma_h$ recursively at every time point using the standard error of the residuals from a univariate autoregressive model for each of the 24 series. The ratio $\frac{\sigma_i}{\sigma_h}$ accounts for the different variability of the hourly price series. A more volatile hour will be assigned a smaller prior variance, effectively keeping coefficients of cross lags shrinkage constant across the different hours. The $\lambda$'s in (8) are hyperparameters, controlling for the amount of shrinkage applied. The exact choice for these values depends on the application at hand. In this paper we simply follow the standard choices as in Koop and Korobilis (2010) and set $\lambda_1 = \lambda_2 = 0.5$ and $\lambda_3 = 100$. This means we do not shrink the coefficients of exogenous variables, so their estimated coefficients are equal to the least squares estimates from the UVAR model.

Given these choices for the prior mean and prior covariance matrix, the posterior for $\boldsymbol{\alpha}$ is given by:

$$
\boldsymbol{\alpha} \mid \boldsymbol{y} \sim N(\boldsymbol{\alpha}^{post}, \boldsymbol{V}^{post}),
\tag{9}
$$

with

$$
\boldsymbol{V}^{post} = \{ (\boldsymbol{V}^{prior})^{-1} + \widehat{\Sigma}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}) \}^{-1},
\tag{10}
$$

$$
\boldsymbol{\alpha}^{post} = \boldsymbol{V}^{post} \{ (\boldsymbol{V}^{prior})^{-1} \boldsymbol{\alpha}^{prior} + (\widehat{\Sigma}^{-1} \otimes \boldsymbol{X})' \boldsymbol{y} \}.
\tag{11}
$$

It is easy to see why the Minnesota prior is a popular choice. First, the posterior and predictive results are available analytically, which greatly facilitates their computation especially in a recursive forecasting exercise as we consider here. Second, there are many adjustments we can apply, including the choice of prior mean vector, the choice of hyperparameters and even the choice of the shrinkage structure. We use the exponentially declining weights as in the original proposal of Doan *et al.* (1984), but alternatively we can treat the exponent as an additional hyper-parameter and optimize it using the in-sample period. For example, Kadiyala and Karlsson (1997) use linearly declining weights, while Carriero *et al.* (2011) perform a grid search over different combinations of hyperparameters. We examined other methods for setting the $\lambda$'s such as optimization and grid

search, but found no evidence for improvement.

### 3.2.3  Factor models

Another common way to account for the curse of dimensionality is through dimension reduction techniques such as factor models (FM) advocated, among others, by Stock and Watson (2002) in the context of macroeconomic forecasting. The general idea of factor models is to summarise the variability in the data using a limited number of $G$, say, linear combinations of the original variables. Stock and Watson (2002) establish the theoretical basis for a two-step forecasting procedure, where in the first step we extract the time series of the $G$ factors $\{\widehat{F}_t\}$ from $Y_t$ using principal component analysis (PCA) using singular value decomposition on the correlation matrix, and then forecast the original dependent variables in the second step. Forecasting can be done in two ways. First, we can project each $y_{h,t+1}$ onto the space spanned by $\{\widehat{F}_t\}$ using least squares, i.e. $\widehat{y}_{h,t+1} = \sum_{g=1}^{G} \widehat{\beta}_{gh} \widehat{f}_{gt}$, where $\widehat{\beta}_{gh}$ is the OLS estimate of the marginal effect of the $g$-th factor $f_{gt}$ on $y_{h,t+1}$. Second, we can model the $G$ factors by means of a VAR model, obtain forecasts $\{\widehat{F}_{t+1}\}$ and use these in the regression $\widehat{y}_{h,t+1} = \sum_{g=1}^{G} \widehat{\beta}_{gh} \widehat{f}_{g,t+1}$. This approach is also referred to as a VAR-PCA model. We performed both procedures but since results are similar, we only report the latter to conserve space. For compatibility, the factor VAR model has the same lag structure as in the other models, namely including only the first, second, and one week lag. The factors are extracted using the deseasonalized price series, and the forecasts are adjusted accordingly. More formally, we estimate the VAR model for the factors, given by

$$\widehat{F}_t = \zeta_1 \widehat{F}_{t-1} + \zeta_2 \widehat{F}_{t-2} + \zeta_3 \widehat{F}_{t-7} + \eta_t, \tag{12}$$

using observations for $t = 1, \ldots, R$. From this model a forecast $\widehat{F}_{R+1|R}$ is obtained, which is used to construct a forecast for the hourly prices as

$$\widehat{Y}_{R+1|R} = \Theta \widehat{F}_{R+1|R} + \Gamma D_{R+1}, \tag{13}$$

where the coefficients $\Gamma$ and $\Theta$ are estimated using information up to time $R$ by regressing the raw hourly price series $Y_R$ on the seasonal dummies $D_R$ and the extracted factors at time $R$.

### 3.2.4  Reduced rank regression

While Principal Component Analysis forms the set of orthogonal latent variables from a subspace spanned by the explanatory matrix $\boldsymbol{X}$, an alternative is to reduce the dimension looking at the subspace spanned by the orthogonal projection of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Reduced rank regression (RRR) does

just that. This technique has a long history in time series analysis (Velu and Reinsel, 1998). Carriero *et al.* (2011) prove consistency and provide the rate of convergence for the estimates when the number of explanatory variables in the system tends to infinity. The basic idea of this approach is to impose a rank restriction on $\Phi$, the matrix of coefficients in (4), and by that focus on a smaller number of underlying components. The reduced rank effectively is imposed by employing the Eckart-Young theorem. Say $\widehat{\boldsymbol{Y}}$ is the matrix of fitted values given by the standard OLS solution that minimizes the error matrix in (4), and let $U\Lambda V'$ be its singular value decomposition, where $\Lambda$ is a diagonal matrix with the singular values arranged in decreasing order $\lambda_1 \geq \cdots \geq \lambda_H$ on its diagonal. We can now cast $\widehat{\boldsymbol{Y}}$ onto a subspace $\widehat{\boldsymbol{Y}}^s = U\Lambda_{(s<H)}V'$, where $\Lambda_{(s<H)}$ equals $\Lambda$ with the last $H-s$ elements on the diagonal set to zero. We can see that in contrast to PCA, RRR pays more attention to the output matrix $\boldsymbol{Y}$ than to the input matrix $\boldsymbol{X}$. Define $\boldsymbol{\Upsilon}_s = \sum_{i=1}^s \nu_i \nu_i'$, where $\nu_i$ is the $i$-th right singular vector from the singular value decomposition of $\widehat{\boldsymbol{Y}}$. The constrained coefficient matrix and the new fitted values are then given by

$$\widehat{\Phi}^{(s)} = \widehat{\Phi}\boldsymbol{\Upsilon}_s, \tag{14}$$

$$\widehat{Y}_t^{(s)} = \widehat{\Phi}^{(s)}X_t. \tag{15}$$

For more details on this procedure, see Izenman (2008).

### 3.2.5 Reduced Rank Bayesian VAR (RRP)

So far we have outlined models that try to avoid the overfitting problem either via shrinkage of the parameters (BVAR) or via dimension reduction (FM and RRR). A method that combines these two approaches is the Reduced Rank Baysian VAR, suggested by Carriero *et al.* (2011). A Reduced rank Bayesian VAR, or Reduced rank posterior (RRP), essentially applies a rank reduction on the posterior estimates obtained from the BVAR. The implementation is similar to the RRR, but instead of the right singular vector $\nu_i$ from the singular value decomposition of $\widehat{\boldsymbol{Y}}$ that was obtained via the UVAR model, we now use $\widehat{\boldsymbol{Y}}$ obtained using the BVAR model. Consequently, our RRP estimator is

$$\widehat{\boldsymbol{\Phi}}_s^{RRP} = \widehat{\boldsymbol{\Phi}}^{BVAR}\boldsymbol{\Upsilon}_s \tag{16}$$

where here $\boldsymbol{\Upsilon}_s = \sum_{i=1}^s \nu_i \nu_i'$ as before, but $\nu_i$ now is the $i$-th right singular vector from the singular value decomposition of $\widehat{\boldsymbol{Y}}^{BVAR}$, and $\widehat{\boldsymbol{\Phi}}^{BVAR}$ is the posterior mean estimate of the BVAR model

coefficients.[5]

## 3.3   Forecast combination

The forecasting performance of different approaches may vary both over time and across different time series, a point nicely illustrated in the context of electricity price forecasting in Weron and Misiorek (2008). There is no apparent reason to restrict ourselves to one method or another. It is now well established that averaging forecasts of different models may very well perform 'better than the best', see Timmermann (2006) for an extensive review.

In the context of forecasting the daily average electricity price, there are two possibilities to form a combined forecast. First, we can directly combine the daily average price forecasts resulting from the different models. For the second possibility, recall that the daily average price is defined as the simple average across the 24 hours. Thus, we may first apply forecast combination weights on the individual hourly price forecasts to form a combined forecast for each hour, and then average those to obtain the daily average price forecast. In the specific case of equal weights for each model for all hours of the day, the two options are equivalent. This is, however, not the case for alternative weighting schemes. The second option obviously is more flexible as the weights assigned to each model are allowed to vary across the individual hours of the day. This need not necessarily result in more accurate forecasts. In general, the forecast combination weights are unknown and need to be estimated. The first option, that is, averaging of daily average price forecasts, involves only a single set of weights; in contrast to the second possibility containing 24 such sets. The increased parameter uncertainty may actually result in worse forecasts for the second option compared to the first. For completeness, we report results from using both combination possibilities.

We examine the performance of two popular ways for forecast combination. We describe the combination schemes in detail for the first option mentioned above, that is, combining daily average price forecasts resulting from different forecast methods. The first combination scheme is the simple average ($AVE$), i.e.,

$$\widehat{\overline{y}}_{t+1|t,AVE} = \frac{1}{W} \sum_{w=1}^{W} \widehat{\overline{y}}_{t+1|t,w}, \tag{17}$$

where $W$ is the number of forecast methods used, and $\widehat{\overline{y}}_{t+1|t,w}$ is the forecast of the daily average price obtained from method $w$.

---

[5]Another closely related model is the Bayesian Reduced Rank Regression introduced by Geweke (1996). A drawback of this model is that it is computationally challenging. Estimation requires simulation involving high dimensional matrix inversion, and can be even more cumbersome in our case as we perform a recursive forecasting exercise. Moreover, Carriero *et al.* (2011) report similar forecasting performance of the two approaches.

Another way to combine forecasts is by estimating weights $a_w$ using a simple linear regression (Bates and Granger, 1969), i.e.,

$$\widehat{\overline{y}}_{t+1|t,LS} = \sum_{w=1}^{W} a_w \widehat{\overline{y}}_{t+1|t,w}. \tag{18}$$

This approach has some drawbacks, however. First, we lose the interpretation of the coefficients $a_w$ as weights, as they can take any value. Negative values or positive values exceeding one are not uncommon and are difficult to interpret in this context. Second, given that different individual forecasts are likely to be highly correlated, there is a possible multicollinearity issue. Hence, the coefficient estimates may vary drastically taking extreme positive or negative values and with very large standard errors, a phenomenon sometimes termed as "bouncing beta's". We therefore pursue a more stable forecast combination approach, by adopting constrained least squares (CLS).

Specifically, at every time point $t$ we numerically solve:

$$\min_{a_w} \frac{1}{Q} \sum_{j=1}^{Q} (\overline{y}_{t+1-j} - \widehat{\overline{y}}_{t+1-j|t-j,CLS})^2, \tag{19}$$

s.t.

$$\widehat{\overline{y}}_{t+1-j|t-j,CLS} = \sum_{w=1}^{W} a_w \widehat{\overline{y}}_{t+1-j|t-j,w}, \tag{20}$$

$$\sum_{w=1}^{W} a_w = 1, \tag{21}$$

$$a_w \geq 0, \qquad w = 1, \ldots, W. \tag{22}$$

In words, we find the forecast combination that minimizes the mean squared forecast error over the most recent $Q$ periods, but restricted such that the weights for the individual forecasts are restricted to be positive and sum to one. We set $Q = 365$, that is, we use a rolling window of one year to estimate the combination weights. We label this approach CLS(A).[6] As mentioned before, another possibility is to first create combined forecasts for each individual hour and then average those to obtain the daily average price forecast. For this purpose, we apply the same CLS procedure to each of the 24 hourly price forecasts to obtain forecast combination weights, which now may vary across the different hours of the day. We label this approach as CLS.

---

[6]We also performed a simple OLS averaging and the inverse of the mean squared forecast error (Stock and Watson, 1998) , i.e. $a_w = \frac{1/MSFE_w}{\sum_{w=1}^{W} 1/MSFE_w}$. Not reported but results from the simple linear regression averaging are poor, and the results for inverse of the mean squared forecast error are similar to the simple average scheme.

# 4   Empirical forecasting results

In this section we present and discuss the results of the empirical forecasting exercise for the daily average electricity price. Our sample includes the last decade, in which electricity markets underwent a process that may have transpired a change in price dynamics. As mentioned in Pesaran *et al.* (2006), a popular way to deal with such changes is by using a rolling window scheme. We estimate all models using a rolling window of five years. This is sufficiently long to provide accurate estimates, yet short enough to allow for parameter instability. The window size is fixed across all models so that we can compare the results using the Giacomini and White (2006) test for unconditional predictive ability.

First we briefly describe the evaluation criteria that we use to assess the accuracy of the forecasts. This includes the test for unconditional predictive ability that we use to directly compare the forecast accuracy for different models. Next we describe the results, where the main finding is that the intraday hourly profile of electricity prices contains valuable information for forecasting the daily average price. We conclude this section with an analysis of the stability of the forecasting performance over time.

## 4.1   Forecasting performance evaluation

We focus mostly on the performance of the different models described in the previous section for one-step ahead point forecasts of the daily average price. For the multivariate models for the panel of the 24 daily hours, these are obtained by averaging the point forecasts for the individual hours, that is

$$\widehat{\overline{y}}_{t+1|t} = \frac{1}{H} \sum_{h=1}^{H} \widehat{y}_{t+1|t,h},$$

where $\widehat{\overline{y}}_{t+1|t}$ is the one-step ahead point forecast for the daily average price on day $t+1$, and $\widehat{y}_{t+1|t,h}$, $h = 1, \ldots, H = 24$, are the individual hourly price forecasts. As a by-product, our multivariate models produce forecasts for the individual hourly prices as well. These forecasts are useful to market participants in their own right as they can also trade electricity for specific hours. Improved hourly forecasts can again be exploited in more efficient trading and bidding strategies for hourly contracts. Hence, in addition we consider the forecasting performance of each model with regards to the individual hours. It is sensible to assume that the best forecasting model for the individual hours will also perform best for the daily average price. That is, if model $A$ performs better than model $B$ for each individual hourly price series, model $A$ is likely to perform better than

model $B$ for forecasting the daily average price. Yet this is not necessarily the case, especially if we do not account for the intraday variation in the variance of the individual hourly prices. A model may perform very well for hours with relatively low volatility but fail for highly volatile hours such that the average across all hours is a poor forecast for the daily average price.

We evaluate the forecast accuracy by means of the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percent error (MAPE), computed as

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=R}^{T-1}(\widehat{\overline{y}}_{t+1|t} - \overline{y}_{t+1})^2}, \tag{23}$$

$$MAE = \frac{1}{N}\sum_{t=R}^{T-1}|\widehat{\overline{y}}_{t+1|t} - \overline{y}_{t+1}|, \tag{24}$$

$$MAPE = \frac{1}{N}\sum_{t=R}^{T-1}\frac{|\widehat{\overline{y}}_{t+1|t} - \overline{y}_{t+1}|}{|\overline{y}_{t+1}|}. \tag{25}$$

where $T$ is the total number of observations, $R$ is the length of the estimation window, and $N = T - R$ is the number of forecasts made. The MAPE evaluates the forecast error relative to the actual price level, and is traditionally used to measure accuracy in electricity load forecasting (Taylor *et al.*, 2006).

When evaluating the accuracy of the hourly price forecasts, we also include a Weighted Root Mean Squared Error (WRMSE), see for example Christoffersen and Diebold (1998). Some hours are more volatile than others, and therefore are harder to predict. When we evaluate the *overall* accuracy of a model with respect to its individual hourly forecasts, it is reasonable to weight the series according to their volatility, so that the more volatile hours will not dominate the evaluation. The $WRMSE$ is calculated as $RMSE'_h\vartheta$ where $RMSE_h$ is a $(24\times1)$ vector of the $RMSE$ measure given above but for the individual hours, and $\vartheta$ is a $(24\times1)$ vector with $(\sigma(y_{t+1,h})/\sum_{h=1}^{H}\sigma(y_{t+1,h}))^{-1}$ at its $h$-th entry, where $\sigma(y_{t+1,h})$ denotes the standard deviation of the $h$-th hourly price.

We address the question of whether the difference in forecasting performance between the models is significant by means of the test for unconditional predictive ability of Giacomini and White (2006). The computation of the test statistic is identical to the test for predictive accuracy in Diebold and Mariano (1995). However, Giacomini and White (2006) generalize the test and develop the theoretical basis for a comparison between *forecast methods*, as opposed to *models*. As a result we can compare between nested and non-nested models, and allow for parameter estimation uncertainty in the forecast evaluation. Assume that we aim to compare two point forecasts $\widehat{\overline{y}}_{t+1|t}^A$ and $\widehat{\overline{y}}_{t+1|t}^B$, obtained with forecast methods $A$ and $B$. We may compare these directly based on their average

accuracy measures (or 'loss functions'), by testing formally whether their difference is statistically significantly different from zero on average. Define the difference in accuracy

$$d_{t+1} = L(\widehat{\overline{y}}_{t+1|t}^A; \overline{y}_{t+1}) - L(\widehat{\overline{y}}_{t+1|t}^B; \overline{y}_{t+1}),$$

where the accuracy measure $L$ can be the squared forecast error $(\widehat{\overline{y}}_{t+1|t} - \overline{y}_{t+1})^2$, the absolute forecast error $|\widehat{\overline{y}}_{t+1|t} - \overline{y}_{t+1})|$, or the absolute percentage error $|\widehat{\overline{y}}_{t+1|t} - \overline{y}_{t+1}|/|\overline{y}_{t+1}|$, corresponding with testing the significance of differences in RMSE, MAE and MAPE, respectively. The null hypothesis of equal predictive ability is given by

$$H_0: \quad \mathsf{E}(d_{t+1}) = 0, \qquad \text{for all } t = R, R+1, \dots, T-1.$$

Let $\overline{d}_{R,N}$ denote the sample average of the differences in accuracy, that is, $\overline{d}_{R,N} = N^{-1} \sum_{t=R}^{T-1} d_{t+1}$. To test the null, we may use a Diebold and Mariano (1995) type statistic

$$t_{R,N} = \frac{\overline{d}_{R,N}}{\sqrt{\hat{\sigma}_{R,N}^2/N}}, \tag{26}$$

where $\hat{\sigma}_{R,N}^2$ is a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator of $\sigma_{R,N}^2 = \mathsf{Var}\left(\sqrt{N}\,\overline{d}_{R,N}\right)$. Under suitable regularity conditions, the statistic $t_{R,N}$ is asymptotically (as $N \to \infty$ with $R$ fixed) standard normally distributed under the null hypothesis.

## 4.2 Results

The forecasting performance for the daily average price is shown in Table 2. The first row gives the values of the RMSE, MAE and MAPE for the AR benchmark model. The performance of the other models is presented in relative terms, in the sense that we show the ratio of the accuracy measure for the specific model over the corresponding measure for the AR model. Hence, a value below one indicates that the specific model provides more accurate forecasts than the AR benchmark.

We observe that, with only a few exceptions, all four types of multivariate models RRR, FM, VAR and RRP perform better than the AR benchmark on all three evaluation criteria. The best individual model is the BVAR model, with improvements of 11%, 17% and 16% in terms of RMSE, MAE and MAPE, respectively. The superior performance of the BVAR relative to the RRR, RRP and FM methods suggests that, from our forecasting perspective, shrinkage is a more useful technique to limit the model complexity than explicit dimension reduction by imposing a reduced

rank or factor structure on the forecasting model.

The results for the three VAR models also show an interesting pattern. First, the DVAR, that is the collection of univariate AR models, performs worse than the AR model for the daily average price. Hence, forecasting the individual hours in isolation does not help to improve forecasts of the daily average price. Second, taking into account the intraday dependence of the hourly prices, as in the UVAR, does lead to more accurate forecasts, despite the large number of parameters that needs to be estimated. In fact, the performance of the UVAR is comparable with the RRR and FM approaches, except that the FM with multiple factors performs substantially better in terms of RMSE. Third, using shrinkage in the VAR model leads to further gains in forecast accuracy, with the BVAR outperforming the UVAR by quite a margin.

Concerning the factor models, it is interesting to note the considerable improvement in forecast accuracy when moving from one factor to two factors. Apparently, while the second factor only explains a relatively small portion of the total variation in the individual hourly prices, as seen in the PCA results in Section 2, it does contain valuable predictive information. This finding complements our previous analysis and also underscores the contribution of the intraday profile to the out-of-sample performance.

Finally, the results in Table 2 demonstrate the attractiveness of forecast combination also in our context of electricity prices. Taking a simple average of all individual forecasts delivers more accurate forecasts than the best individual model (that is, the BVAR), although the differences are rather small at 1%. More substantial improvements in forecast accuracy can be obtained by allowing for different weights for the individual forecasts using the CLS method. The resulting forecasts provide an improvement of 16%, 20% and 18% in terms of RMSE, MAE and MAPE relative to the AR benchmark. Both options, with weights determined by a CLS for each individual hour or by CLS for the daily average price, are better than equal weighting of the different models. Allowing the weights to vary across individual hours produces slightly better results. We leave a more detailed comparison between the two approaches for future research.

Given the superior performance of the CLS forecast combination method, it is interesting to examine which models receive most weight in this approach, and whether and how the weights vary over time For this purpose, Figure 5 shows the time series of estimated weights, averaged across hours, obtained with the CLS procedure. The RRP and FM weights plotted are the sum for these models with different ranks and number of factors considered (that is, 1, 2, and 5). For convenience, the daily average price process is plotted as well.

Several features stand out from the graph. First, the weights are not constant over time.

|  | RMSE | MAE | MAPE |
|---|---|---|---|
| *Univariate models* | | | |
| AR* | 23.41 | 11.93 | 0.054 |
| HAR | 1.33 | 1.44 | 1.432 |

\* Actual terms, not relative.

|  | RMSE | MAE | MAPE |
|---|---|---|---|
| *VAR models* | | | |
| DVAR | 1.00 | 1.03 | 1.02 |
| UVAR | 0.97 | 0.88 | 0.89 |
| BVAR | 0.89 | 0.83 | 0.84 |
| *Factor models* | | | |
| FM(1) | 0.96 | 0.93 | 0.92 |
| FM(2) | 0.91 | 0.91 | 0.90 |
| FM(5) | 0.90 | 0.88 | 0.89 |
| *Reduced Rank models* | | | |
| RRR(1) | 1.01 | 0.89 | 0.90 |
| RRR(2) | 0.98 | 0.88 | 0.89 |
| RRR(5) | 0.98 | 0.85 | 0.89 |
| RRP(1) | 0.92 | 0.91 | 0.96 |
| RRP(2) | 0.91 | 0.90 | 0.95 |
| RRP(5) | 0.91 | 0.90 | 0.95 |
| *Forecast Combination* | | | |
| AVE | 0.88 | 0.82 | 0.83 |
| CLS | **0.84** | **0.80** | **0.82** |
| CLS(A) | **0.85** | **0.82** | **0.83** |

**Table 2:** Performance comparison between the different models, for the daily average price. The first row gives the values of the RMSE, MAE and MAPE for the AR benchmark. The performance of the other models are presented in relative terms, in the sense that we show the ratio of the accuracy measure for the specific model over the corresponding measure for the AR model. HAR: Heterogeneous Autoregressive model, DVAR: Diagonal VAR, UVAR: Unrestricted VAR, BVAR: Bayesian VAR, FM is factor model with the number of factors in parentheses, RRR: Reduced Rank Regression with rank in parentheses, RRP: Reduced Rank Posterior with rank in parentheses. For Forecast Combination, AVE stands for simple averaging, CLS stands for Constrained Least Squares weights with weights that may vary across hours, CLS(A) stands for Constrained Least Squares weights with weights which are determined using the daily average price forecast. The forecasting exercise is performed using a rolling estimation window of five years. The forecast period runs from May 4, 1998 to March 10, 2010 which corresponds with $N = 4694$ forecasts.

For example, the FM models receive a large weight between 0.4-0.5 during the years 1999-2002 but a substantially lower weight between 0.2-0.3 in subsequent years. This suggests that there is considerable variation in relative performance of the different models over time. Second, the weights are quite volatile, despite being determined by a rolling window of a full calendar year. Hence, changes in the relative accuracy of individual forecasts may occur rapidly. Third, the weights of the BVAR and the RRP increase during upwards spikes, but also during volatile periods like the years 2003-2006. Fourth and finally, the weights assigned to the DVAR are fluctuating between 0.2-0.3 in the early years of our sample period, often being ranked second (after the FMs) in terms of relative performance. In the years 2006-2010, the weight given to the DVAR forecasts declines to around 0.1 and is, most of the time, dominated by the other models. Apparently, from our forecasting perspective, incorporating cross-sectional dependence between the individual hourly prices has become more important in recent years.

As discussed earlier in this section, we address the significance of differences in the forecast accuracy measures by means of the Giacomini-White test statistic for unconditional predictive ability. Table 3 presents the test results based on the absolute error loss function, corresponding with testing the null hypothesis of equal predictive ability in terms of MAE.[7] For RRR, RRP and FM, we only include the best individual specification, that is, with rank five for RRR and RRP and with five factors for FM. The statistic is computed such that a positive statistic means that the MAE of the row model is larger than the MAE of the column model.

Table 3 shows that at conventional confidence levels the BVAR significantly outperforms all other individual forecast methods. This includes its unrestricted version, the UVAR, which demonstrates the efficiency and appropriateness of the shrinkage procedure in this case. All individual forecasts also significantly outperform the benchmark AR model. The forecast combination method with weights determined by constrained least squares (CLS) gives significantly more accurate forecasts than all individual models at the 1% significance level, except the BVAR. Based on a one-sided test, the value of the Giacomini-White statistic of $-1.34$ corresponds with a $p$-value of 0.09.

Table 4 presents the forecast accuracy measures for the one-step ahead point forecasts of the individual hourly prices. Apart from the $WRMSE$, these measures are calculated for each hour separately and then simply averaged across the hours. The $WRMSE$ is a weighted average of the $RMSE$ as explained in the previous section.

Several conclusions emerge from Table 4, which mostly corroborate the forecasting results for

---

[7]Results for testing the null hypothesis of equal predictive ability in terms of RMSE and MAPE are qualitatively similar. Details are available upon request.
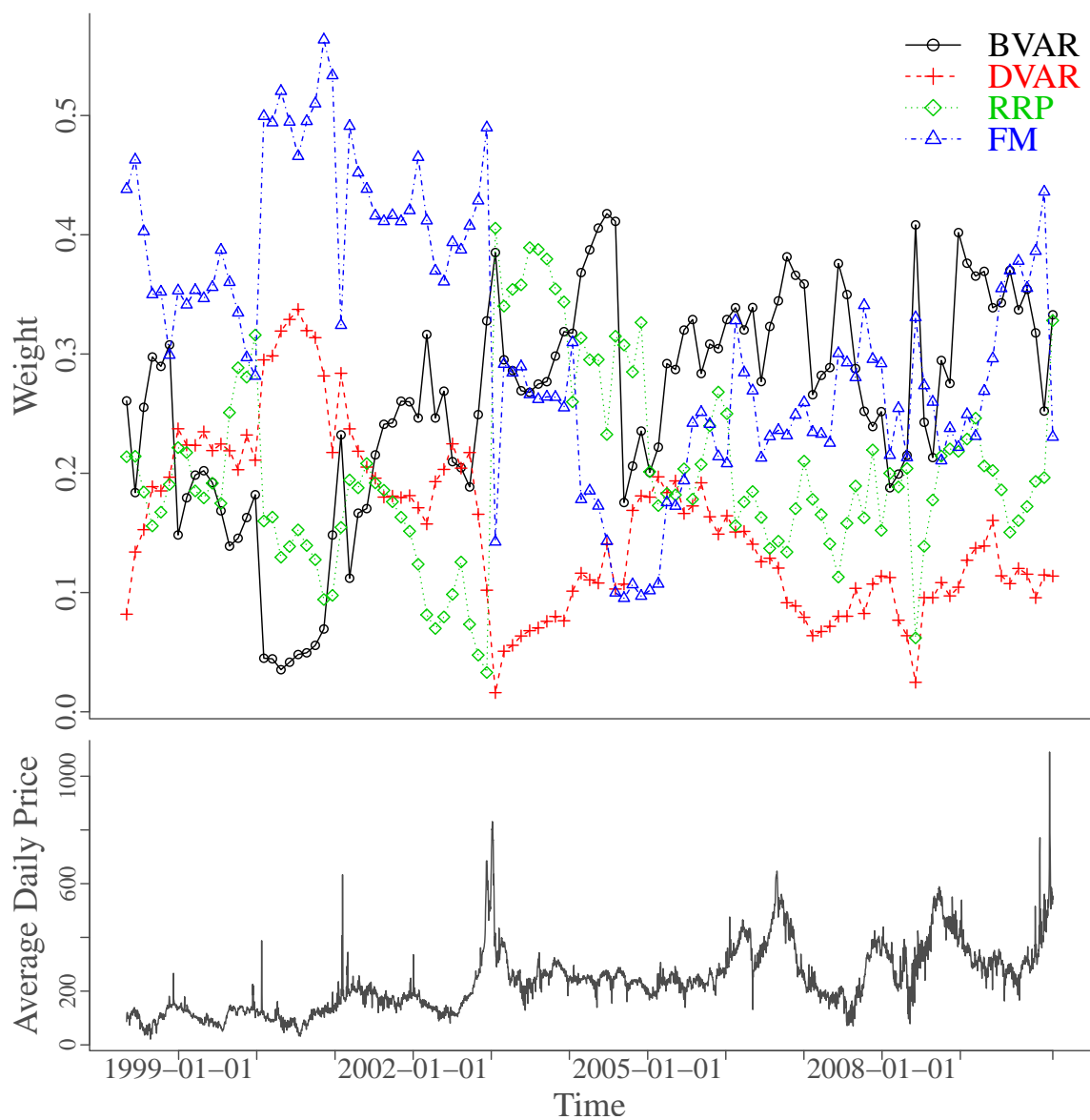
**Figure 5:** Averaged CLS weights across the 24 hours for different types of multivariate models over time. The weights are determined using a constrained least squares procedure over the most recent 365 days in a rolling window scheme. BVAR: Bayesian VAR, DVAR: Diagonal VAR, RRP: Reduced Rank Posterior, FM: Factor Model. The RRP and FM weights plotted are the sum for these models with different ranks and number of factors considered (that is, 1, 2, and 5). The bottom panel presents the daily average price over time.

|       | BVAR  | UVAR  | DVAR  | RRP(5) | FM(5) | RRR(5) | CLS  |
|-------|-------|-------|-------|--------|-------|--------|------|
| UVAR  | 5.52  |       |       |        |       |        |      |
| DVAR  | 7.90  | 5.20  |       |        |       |        |      |
| RRP(5)| 10.47 | 1.73  | -4.40 |        |       |        |      |
| FM(5) | 2.50  | -0.72 | -7.98 | -2.33  |       |        |      |
| RRR(5)| 5.55  | 3.14  | -5.20 | -1.71  | 0.73  |        |      |
| CLS   | -1.34 | -5.50 | -9.02 | -9.92  | -3.35 | -5.53  |      |
| AR    | 6.22  | 4.09  | -1.93 | 3.18   | 5.58  | 4.08   | 6.96 |

**Table 3:** Giacomini-White test statistics based on the absolute forecast error loss function. The statistic is computed such that a positive statistic means that the MAE of the row model is larger than the MAE of the column model.

the daily average price discussed before. First, across all individual models the BVAR renders the most accurate forecasts for the hourly electricity prices, with improvements relative to the DVAR model between 7-13% depending on which accuracy measure is used. Hence, it comes as no surprise that the BVAR turns out to be performing best for the daily average price in Table 2. Second, for the factor models it again seems important to include multiple factors, in the sense that the FM(1) forecasts are substantially less accurate (by 10% or more) compared to the FM(2) and FM(5) forecasts. Interestingly, the same conclusion applies to the RRR and RRP forecasts, where the models with higher rank (2 and 5) outperform the models with rank equal to one. Third, forecast combination also gives superior forecasts for the individual hourly prices, with the CLS method improving upon the BVAR approach by another 2-4%. Note, however, that in this case simply averaging all individual forecasts is not sufficient to reap the gains of forecast combination. In fact, the simple average forecast combination (AVE) performs substantially worse than the BVAR as well as several other individual forecasts. Finally, in Table 2 the DVAR model is seen to give the least accurate forecasts of the daily average price across all individual models. This no longer holds for the forecasts of the individual hourly prices. As seen in Table 4, the DVAR outperforms the RRR(1) and RRP(1) forecasts as well as the FM(1) forecasts. This can partly be explained by these approaches rendering (relatively) more accurate forecasts for relatively volatile hours of the day, as suggested by the relatively large values of the WRMSE compared to the (unweighted) RMSE.

### 4.3 Stability Analysis

The analysis so far suggests that there are significant improvements in forecast accuracy for the daily average price to be gained by opting for a multivariate framework using the full panel of hourly electricity prices. We now examine how robust this conclusion is with respect to the forecast period.

|  | *WRMSE* | *RMSE* | *MAE* | *MAPE* |
|---|---|---|---|---|
| *VAR models* | | | | |
| DVAR* | 28.45 | 30.62 | 14.03 | 0.07 |
| UVAR | 0.96 | 0.97 | 0.89 | 0.91 |
| BVAR | 0.93 | 0.93 | 0.88 | 0.87 |
| * Actual terms, not relative. | | | | |
| | | | | |
| *Factor models* | | | | |
| FM(1) | 1.08 | 1.05 | 1.09 | 1.10 |
| FM(2) | 0.97 | 0.96 | 0.97 | 0.97 |
| FM(5) | 0.94 | 0.94 | 0.92 | 0.93 |
| | | | | |
| *Reduced Rank models* | | | | |
| RRR(1) | 1.16 | 1.13 | 1.15 | 1.19 |
| RRR(2) | 1.01 | 1.01 | 0.99 | 1.03 |
| RRR(5) | 0.96 | 0.97 | 0.91 | 0.93 |
| RRP(1) | 1.11 | 1.08 | 1.15 | 1.22 |
| RRP(2) | 1.00 | 0.99 | 1.05 | 1.07 |
| RRP(5) | 0.95 | 0.95 | 0.98 | 0.99 |
| | | | | |
| *Forecast Combination* | | | | |
| AVE | 0.98 | 0.97 | 0.94 | 0.97 |
| CLS | **0.89** | **0.89** | **0.85** | **0.85** |

**Table 4:** Performance comparison between the different models. The performance measures are calculated for each individual hour and averaged across the 24 hours. The first row gives the values of the WRMSE, RMSE, MAE and MAPE for the Diagonal VAR (DVAR). The performance of the other models are presented in relative terms, in the sense that we show the ratio of the accuracy measure for the specific model over the corresponding measure for the DVAR model. UVAR stands for Unrestricted VAR, BVAR: Bayesian VAR, FM is factor model with the number of factors in parentheses, RRR: Reduced Rank Regression with rank in parentheses, RRP: Reduced Rank Posterior with rank in parentheses, AVE stands for simple averaging and CLS stands for Constrained Least Squares weights, both for the Forecast Combination method. The forecasting exercise is performed using a rolling window of five years. The forecast period runs from May 4, 1998 to March 10, 2010.

Figure 6 presents a rolling MAPE ratio of selected models relative to the benchmark AR model, computed using a moving window of three years. We observe that both the BVAR framework as well as the CLS forecast combinations deliver consistent and substantial improvements in forecast accuracy relative to the benchmark. Of course there is variation in the relative accuracy. For example, the relative MAPE for the CLS approach varies between 0.88 for windows covering the period 1999-2001 and 0.73 for the years 2003-2005. The forecast gains always exceed 10%, however. The same conclusion applies to the FM(5) method, although the gains in MAPE are smaller, ranging between 5-15%, depending on the window considered. For the other forecast methods, the forecast accuracy shows more instability over time. For RRP(5), the positive full-sample results presented in Table 2 appear to be mostly due to the final years of the forecast period, in the sense that the relative MAPE hovers around one until 2007 and only then declines to substantially lower values. Similarly, the unrestricted VAR forecasts do not improve upon the benchmark in the first years of the forecast period. They do show substantial improvement as time goes by though, even outperforming the FM(5) model from 2004 onwards. Finally, the Diagonal VAR, which is restricted from using the information in the cross-sectional relations of the hourly prices, performs slightly worse than the AR benchmark throughout the forecast period.

# 5   Conclusion

In liberalized electricity markets such as Nord Pool, the daily electricity price is an average of the set of prices for delivery during individual hours of the day, which are determined simultaneously in a day-ahead auction market. In this paper, we present convincing empirical evidence that, for the purpose of forecasting the daily average electricity price, it is beneficial to exploit the information embedded in the panel of the hourly price series. This can be done by adopting a multidimensional modeling framework for the individual hourly prices. A key requirement to realize the improvements in forecast accuracy is incorporating the complex intraday relations between the hourly prices. A collection of univariate autoregressive models for the individual hours does not outperform a univariate AR benchmark for the daily average price. However, allowing for cross-sectional effects substantially improves performance. Using dimension reduction techniques and, in particular, shrinkage and forecast combination further improve out-of-sample performance, resulting in a significant improvement in forecast accuracy of about 15-20% compared with a univariate forecast method for the daily average price itself.
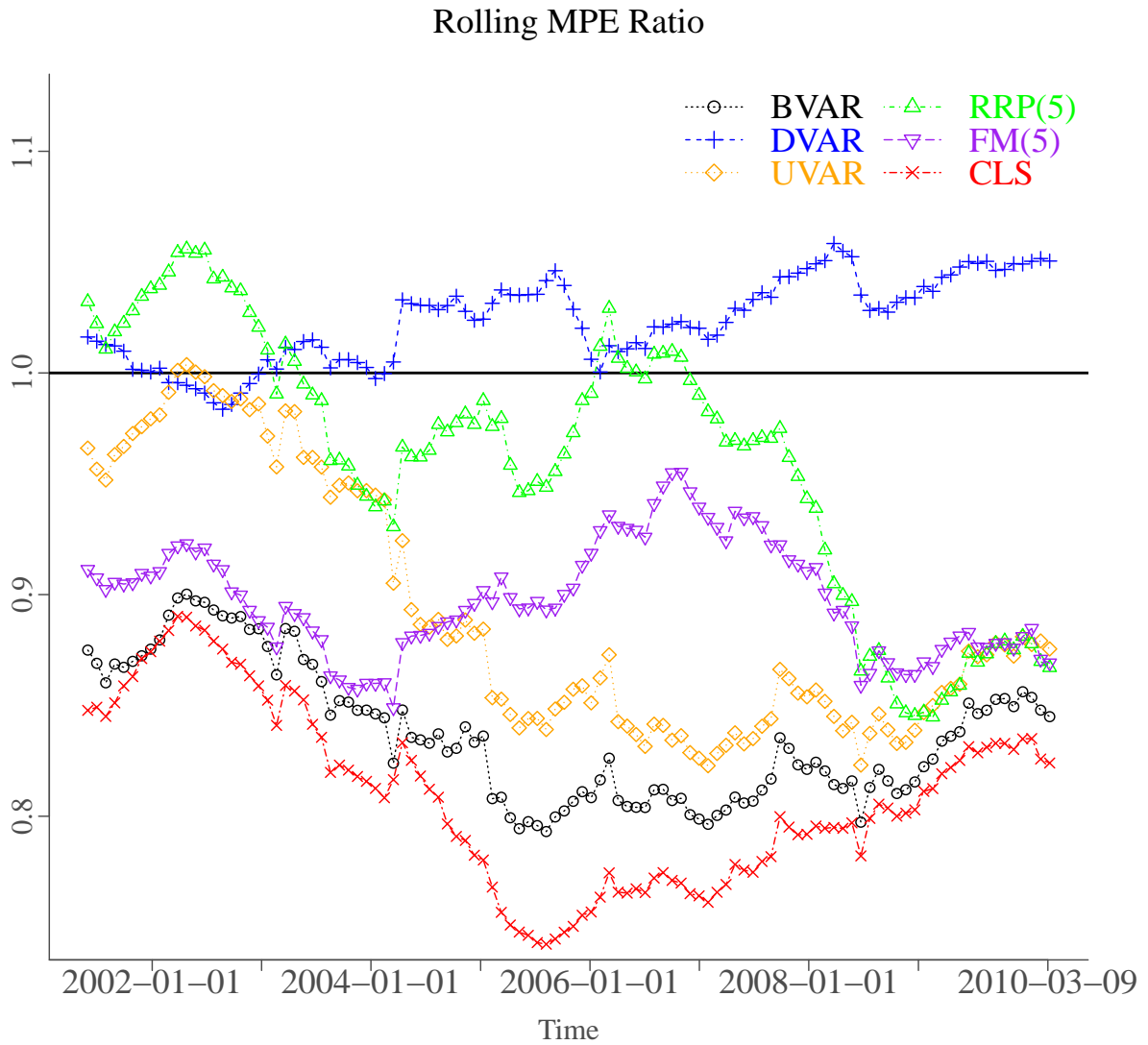
**Figure 6:** Three years rolling MAPE measure. The graph shows the ratio of the MAPE of selected models and the AR benchmark.

# References

Bates, J. M. and C. W. J. Granger (1969), The combination of forecasts, *Operations Research Quarterly*, 20, 451–468.

Benth, F. E. and S. Koekebakker (2008), Stochastic modeling of financial electricity contracts, *Energy Economics*, 30, 1116–1157.

Boogert, A. and D. Dupont (2008), When supply meets demand: The case of hourly spot electricity prices, *IEEE Transactions on Power Systems*, 23, 389–398.

Carriero, A., G. Kapetanios, and M. Marcellino (2011), Forecasting large datasets with Bayesian reduced rank multivariate models, *Journal of Applied Econometrics*, 26, 735–761.

Christensen, T. M., A. S. Hurn, and K. A. Lindsay (2012), Forecasting spikes in electricity prices, *International Journal of Forecasting*, 28, 400–411.

Christoffersen, P. and F. X. Diebold (1998), Cointegration and long-horizon forecasting, *Journal of Business and Economic Statistics*, 16, 450–458.

Corsi, F. (2009), A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, 7, 174–196.

Cuaresma, J. C., J. Hlouskova, S. Kossmeier, and M. Obersteiner (2004), Forecasting electricity spot-prices using linear univariate time-series models, *Applied Energy*, 77, 87–106.

Diebold, F. X. and R. Mariano (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, 253–265.

Doan, T., R. Litterman, and C. A. Sims (1984), Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews*, 3, 1–144.

Geweke, J. (1996), Bayesian reduced rank regression in econometrics, *Journal of Econometrics*, 75, 121–146.

Giacomini, R. and H. White (2006), Tests of Conditional Predictive Ability, *Econometrica*, 74, No. 6, 1545–1578.

Härdle, W. K. and S. Trück (2010), The dynamics of hourly electricity prices, SFB 649 Discussion Paper 2010-013.

Huisman, R., C. Huurman, and R. Mahieu (2007), Hourly electricity prices in day-ahead markets, *Energy Economics*, 29, 240–248.

Huurman, C., F. Ravazzolo, and C. Zhou (2012), The power of weather, *Computational Statistics & Data Analysis*, 56, 3793–3807.

Izenman, A. J. (2008), *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer.

Kadiyala, K. R. and S. Karlsson (1997), Numerical methods for estimation and inference in Bayesian VAR models, *Journal of Applied Econometrics*, 12, 99–132.

Karakatsani, N. V. and D. W. Bunn (2008), Modelling the volatility of spot electricity prices, *International Journal of Forecasting*, 24, 764–785.

Knittel, C. R. and M. R. Roberts (2005), An empirical examination of restructured electricity prices, *Energy Economics*, 27, 791–817.

Koop, G. and D. Korobilis (2010), Bayesian multivariate time series methods for empirical macroeconomics, *Foundations and Trends in Econometrics*, 3, 267–358.

Kristiansen, T. (2012), Forecasting Nord Pool day-ahead prices with an autoregressive model, *Energy Policy*, 49, 328 – 332.

Longstaff, F. A. and A. W. Wang (2004), Electricity forward prices: A high-frequency empirical analysis, *Journal of Finance*, 59, 1877–1900.

Möst, D. and D. Keles (2010), A survey of stochastic modelling approaches for liberalised electricity markets, *European Journal of Operational Research*, 207, 543–556.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006), Forecasting time series subject to multiple structural breaks, *The Review of Economic Studies*, 73, 1057–1084.

Stock, J. H. and M. W. Watson (2002), Forecasting using principal components from a large number of predictors, *Journal of the American Statististical Association*, 97, 1167 – 1179.

Taylor, J. W., L. M. De Menezes, and P. E. McSharry (2006), A comparison of univariate methods for forecasting electricity demand up to a day ahead, *International Journal of Forecasting*, 22, 1–16.

Timmermann, A. (2006), Forecast combinations, in G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, Elsevier, pp. 135–196.

Velu, R. and G. C. Reinsel (1998), *Multivariate Reduced-Rank Regression*, springer.

Weron, R. (2006), *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, John Wiley & Sons.

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

Zareipour, H. (2012), Short-term electricity market prices: A review of characteristics and forecasting methods, in A. Sorokin, S. Rebennack, P. M. Pardalos, N. A. Iliadis, and M. V. F. Pereira (eds.), *Handbook of Networks in Power Systems I*, Springer, pp. 89–121.