Forecasting short term electricity prices

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

[1] *Econometric Institute, Erasmus University Rotterdam*

[2] *Tinbergen Institute*

May 22, 2012

Abstract

In this paper we forecast the day-ahead electricity spot price. We show that taking into account the intra-day relation between individual hourly prices leads to significant improvement in forecast accuracy. Several multivariate models are estimated, models that allow for cross lags effect, as opposed to just own lags. We compare these models with a viable univariate modelling alternative, a dynamic $ARX(p)$. We deal with the inherent over-fitting problem using shrinkage and dimension reduction methods such as Bayesian VAR, reduced-rank regression and principal components regression. We find that additional gain is achieved using forecast combination, an average reduction of 16% for the RMSE metric, compared with the flexible benchmark model.

[1] Corresponding author, Email: `raviv@ese.eur.nl`

1 Introduction

Electricity prices exhibit several unique features, mainly due to the non-storable nature of this *flow* commodity. Shocks to demand caused by, say, extremely low temperature can not be smoothed out using pre-stored inventory as it happens for other, *stock* commodities. As a direct result electricity prices display extremely high volatility, much higher than other energy products. Other unique characteristics include possible negative prices, weekly and monthly seasonality, sudden large price changes (”spikes”), and a mean reversion effect, see Knittel and Roberts (2005), or Longstaff and Wang (2004) for more details.

The market in which short term electricity contracts, hourly and day ahead contracts, are traded is referred to as a spot market or day-ahead market. In this paper we use data from *Nordpool spot* market. It is a physical market where participants trade electricity for the next day. The quotes are submitted *simultaneously* for all hours of the next day and market prices are determined by the intersection point of the aggregate demand and supply curves and is quoted for delivery for every hour of the next day. As a result, in contrast to the stock market, information about the price for contract that delivers electricity between 20:00-21:00h, cannot be exploited to predict the price of a contract that delivers electricity between 21:00-22:00h, as these two prices are determined simultaneously. Hourly prices should not be modelled as a time series process but as a cross sectional panel that vary from one day to the next. (See Huisman et al., 2007). The spot price itself is defined as the average of the 24 individual hourly prices.

A good prediction of both hourly prices and the spot price itself is important for several reasons. Firstly, the spot price is used as a reference for derivative pricing, e.g. hourly power options or daily callable options. Secondly, companies that use electricity for production might want to trade contracts that deliver electricity for specific hours of the day, rather than the standard Base load contract that covers all 24 hours. Lastly, using improved set of forecasts, market participants may develop efficient bidding strategies that help to control risk and increase profits.

Our objective in this paper is twofold. First, we show that taking into account the intra-day relation between the individual hours is beneficial for spot price forecasting. Second, we show that we can further benefit from a more complex multivariate framework which allows each hourly price to be modelled and predicted separately, the forecast for the spot price is then the average of the 24 individual hourly price forecasts.

In the last couple of decades there has been an ongoing liberalization of electricity markets. This fact has created an increasing interest in building econometric models for electricity prices. Weron and Misiorek (2008) report forecasting results for the spot price from a variety of linear and non-linear models, including basic autoregressive models, jump-diffusion models and regime-switching models. Bunn and Karakatsani (2008) add fundamental variables such as fuel prices and level of demand. **?** implement regression model with seasonal periodic autoregressive fractionally integrated moving average disturbances, but do not report any forecasting results. **?** use data from Leipzig Power Exchange (LPX), they implement univariate time series models such as AR, ARMA and unobserved components, they also allow for jumps and time-varying intercepts. Upon comparison, they conclude that modelling every hour separately produces better forecasting results when compared with a model for the spot price itself. We find that modelling each hour separately is not sufficient to produce superior forecasting results. The intra-day relation between the individual hours is the information source which leads to more accurate forecast of the spot price.

Modelling the individual hours separately increases complexity. In order to extend our model and forecast the spot price using its individual components, i.e. hourly prices, we inflate the number of parameters. A comparison between AR(1) model, say, for the spot price and AR(1) model forecasts averaged across the 24 hours is somewhat unjust. A practitioner who wishes to avoid the curse of dimensionality and forecast the spot price using a univariate model, will not necessarily opt for AR(1) but a more flexible alternative, e.g. AR(7), which is less parsimonious, yet is nonetheless much less parametrized than the multivariate, 24 hourly price forecasts alternative. Hence, we should compare our more complex multivariate model with a flexible alternative model for the univariate spot price, for example, an AR(p) model, as this is a more realistic and viable alternative in practice. Hendry and Hubrich (2006) show that forecasting the aggregate, or the spot price in our case, is theoretically never better than forecasting the aggregate using its disaggregate components, the hourly prices in our case. Nevertheless, their result is complicated in practice by possible parameter instability and measurement error, thus it is not straight forward to argue in favour of forecasting the spot price using its individual components.

A trivial way to forecast the spot price using the 24 daily hours, and also to allow for cross relation between the individual hours, is via VAR-type models with lagged and cross lags coefficients. Such models are highly flexible but far from parsimonious. A VAR model for the 24 hourly prices holds a large number of parameters, and hence high estimation uncertainty, e.g. VAR(3) with no exogenous variables apart from the intercept has $(1 + 3 \times 24) \times 24 = 1752$ parameters that are essential for point forecasting. This effectively dissipates our degrees of freedom, which brings about potential over-fitting. The model will not only fit the data, but also the intrinsic noise in it. The problem worsens for data with such large short term swings as electricity prices, the model is much more likely to overshoot its forecasts due to estimates that are too close to the in-sample behaviour.

We cope with this problem of over fitting in two ways. The first is through dimension reduction techniques such as Reduced Rank Regression and Factor Models. The second is via regularization, or shrinkage. A Bayesian VAR model can be viewed as one such method. We enforce a prior on the parameters, efficiently shrinking parameter estimates to mitigate over-fitting. We also combine these two methods by way of forecasts averaging, in section 4 we report results for simple model averaging, and for constrained OLS averaging. We investigate the forecasting performance of our set of models using data from "Nordpool power exchange", covering hourly electricity prices during [4th May, 1992 - 4th March, 2010]. We compare the accuracy of our models for both the spot price forecasts and the individual hours forecasts. We find that for spot price forecasting a multivariate framework which allows for hourly cross section relation has substantial forecasting gains. These gains can be further improved upon by combining forecasts from different models. The comparison is done with respect to a flexible alternative, a dynamic ARX(p) model for the spot price. On average, we achieve a reduction of 16% for the RMSE evaluation metric. The rest of the paper is organized as follows, Section 2 describes the data, Section 3 introduces the set of models we use for forecasting, Section 4 presents the results, discuss their significance and assess their robustness, we conclude in Section 5.

2 Data Analysis

We use data from the Nordic power exchange, Nordpool, which includes Denmark, Norway, Sweden, Finland, and Germany. Elspot is Nord Pool Spot's day-ahead auction market. It is the largest market for electrical energy in the world. About 330 companies from 20 countries trade on the market and participants include both producers and large consumers. Every day participants submit bids and offers hour by hour, through Nord Pool Spot's web-based trading system. They can place their orders until 12:00 CET. Buy and sell orders are aggregated into demand and supply curves for each delivery hour. The system price for each hour of the next day is then determined by the intersection of these curves. Prices are quoted for megawatt per hour (MWh). Our sample covers twenty-four hourly prices for each day for the period 4th May, 1992 - 4th March, 2010 (6519 days). Since our sample dates back to the pre-euro currency era, prices are quoted in Norwegian krone (NOK). Table 1 presents the descriptive statistics.

We see the high volatility feature typical for electricity prices. Standard deviation is about half the mean price. Excess kurtosis is high for hours (8-10) in the early morning and early evening (18-19), these hours are more likely to exhibit extreme prices.

Hour	Mean	SD	Skewness	Excess Kurtosis
1	194.23	111.40	1.07	1.89
2	188.73	109.34	1.07	1.87
3	185.03	108.48	1.10	2.02
4	182.29	107.09	1.08	1.95
5	183.12	108.54	1.12	2.09
6	189.36	111.52	1.11	2.07
7	199.10	114.84	1.08	1.91
8	211.83	123.76	1.47	5.71
9	221.98	140.12	3.09	30.68
10	221.74	132.29	2.31	19.44
11	221.33	126.88	1.53	6.75
12	219.47	123.52	1.26	3.43
13	215.94	120.78	1.14	2.39
14	213.46	119.08	1.11	2.11
15	211.57	118.16	1.11	2.08
16	210.71	118.42	1.12	2.01
17	212.37	122.95	1.32	3.42
18	217.25	131.10	1.72	7.18
19	216.52	126.81	1.40	4.21
20	212.72	120.84	1.15	2.32
21	209.26	117.25	1.08	1.84
22	207.96	115.85	1.07	1.85
23	204.89	114.01	1.04	1.66
24	197.20	110.49	1.05	1.72
Spot	206.16	116.63	1.13	2.21

Table 1: Descriptive statistics of Nord-Pool hourly electricity prices. Prices are quoted in Norwegian krone (NOK). During the sample period one Euro was approximately 8.5 NOK.

Figure 1 provides a more visual description of our dataset. Spikes are clearly visible in the top left quadrant. Looking at the correlation structure at the top right of figure 1, we can see that night hours vary together while early morning and early evening hours have a more independent nature, e.g. hour number 9 (between 08:00 AM and 09:00 AM) has relatively weak correlation with the rest of the day. Having said that, correlation is

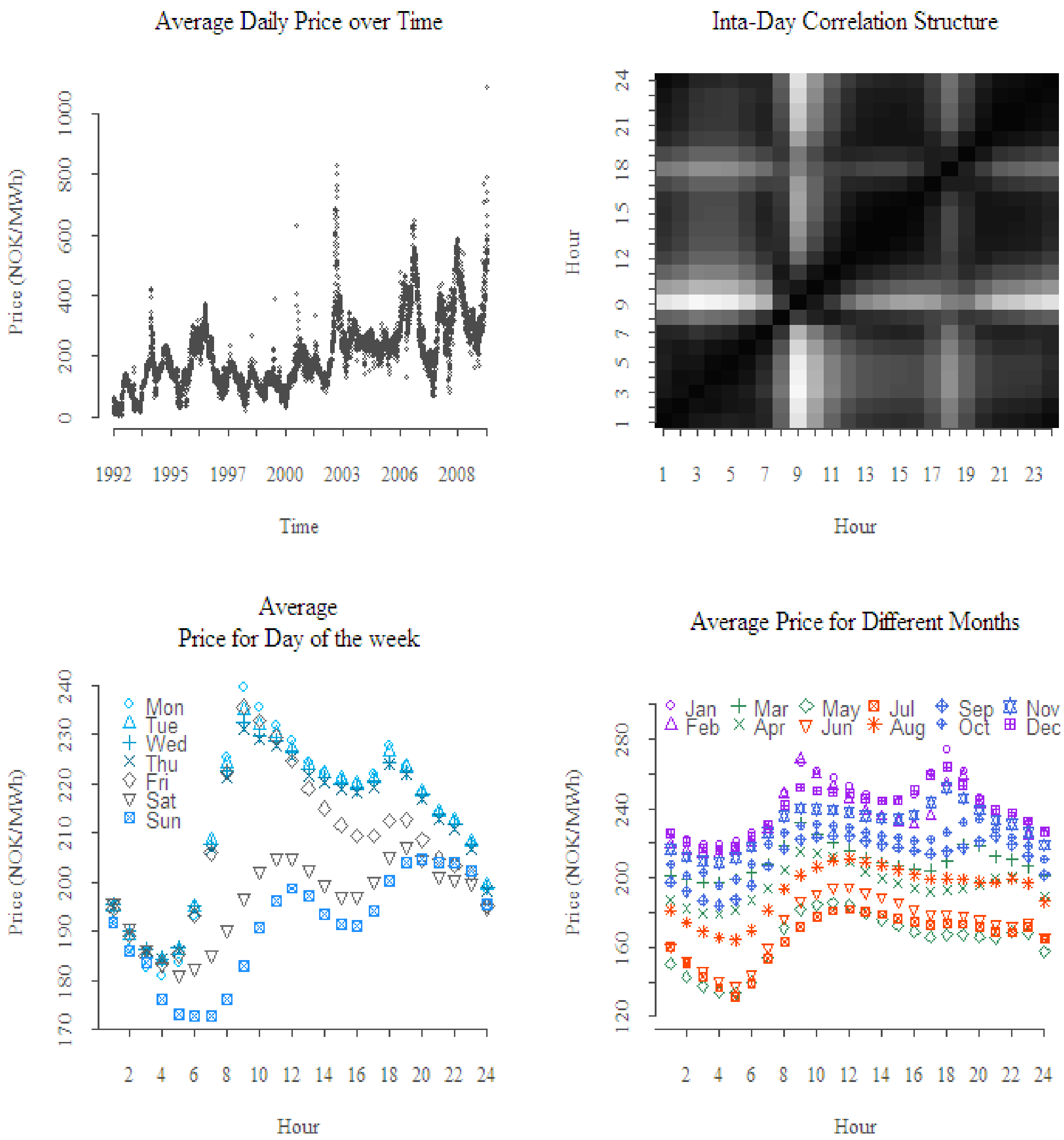

Figure 1: **Top left**: Spot price over time. **Top right**: Hourly correlation structure, the darker the square, the higher the correlation between the hours. We can see for example that hour number 9 has relatively lower correlation with the rest of the hours. **Bottom left**: Average price according across days of the week. **Bottom right**: Average hourly price across months, the different colors represent different seasons.

generally high, between 0.85 and 0.99. The bottom half of the figure displays the other two seasonality components observed for electricity prices. We see that weekdays have generally higher prices than weekends, but also that the two weekend days are slightly different, so one might want to address them separately. In the bottom right we observe that, as expected, winter months experience higher prices than summer months, and that the pattern is quite monotonic, with summer months having the lowest average and winter months having the highest.

2.1 Principal Component Analysis (PCA)

In the previous section we mentioned the high cross correlation between the hourly prices, thus it is reasonable that few common factors drive the bulk of co-movement. In order to get more insight about these determinants we perform Principal Component Analysis using the correlation matrix.[1] Figure **??** presents the first two principal components along with their corresponding loadings. Not surprisingly, the first factor is simply a level factor. We interpret the second factor as the spread between prices during peak hours (8-12, 17-19) and prices during off-peak hours. The first common factor explains about 96.2% of the joint variation, the second component adds another 2.2%. The remaining 22 components account for the last 1.6% of variability in the data. The high percentages of explained variance for the first few factors encourages the use of dimension reduction techniques for forecasting, e.g. Principal Component Regression (PCR). The loadings of the first principal component naturally indicate that this is a level factor. We can interpret the second factor as the spread between prices during peak hours (7:00-11:00 and 18:00-20:00) and off-peak hours.

3 Models and Estimation

In this section we describe the models used for prediction. Mean reversion is a well documented property of short term energy prices, see for example Knittel and Roberts

[1] For more details on these procedure see **?**.

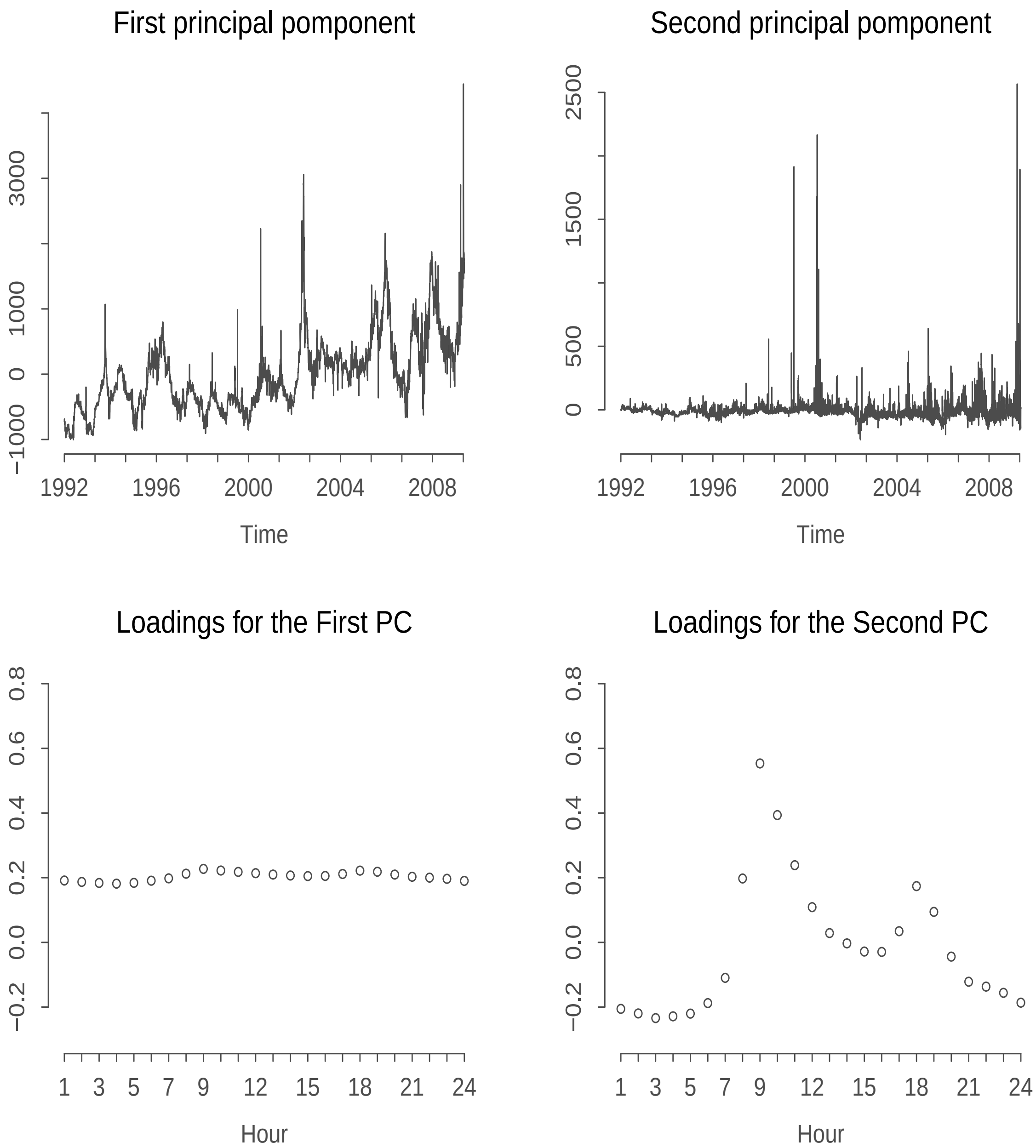

Figure 2: **Top left**: First Principal Component over time. **Top right**: Second Principal Component over time. **Bottom left**: Loadings on the first principal component. **Bottom right**: Loadings on the second principal component.

(2005). Thus, for forecasting purposes, we carry out our analysis working with price levels and avoid any data transformation. We adopt a rolling window scheme for parameter estimation for all models. We do so to account for the potential problem of structural breaks, Pesaran and Timmermann (2005).). For example, dynamics might have changed during the worldwide electricity market deregulation over the last decade.

3.1 Univariate Models

The spot price is a univariate time series defined as a the average price over the 24 individual hourly prices:

$$\bar{y}_t = \frac{\sum_{j=1}^{24} y_{jt}}{24} \tag{1}$$

y_{jt} denotes the price for hour j at day t.

We aim (1) to stress the importance of the information embedded in intra-day relation between hours for forecasting the spot price, and (2) give evidence for the benefits of using a multivariate framework, that is, modelling every hourly price series individually. In order to show that it is indeed preferable to use more complex multivariate models, we want to compare these with a reasonable univariate alternative, a benchmark that may be used in practice, as oppose to a random walk benchmark[2]. We now present two univariate models which capture solely spot price dynamics, and ignore the idiosyncratic information of each individual hour. A Heterogeneous Autoregressive model (HAR) model and a Dynamic $ARX(p)$ model which will also serve as our benchmark model, as it is a sensible alternative to the more complex multivariate models that follow.

3.1.1 Heterogeneous Autoregressive model (HAR)

The Heterogeneous Autoregressive model (HAR) is an AR-type model which was recently used by Corsi (2009) to forecast realized volatility. The model is designed to capture long memory behaviour, also observed in energy prices. The advantage of this model is that it circumvent the need to estimate every lag coefficient separately, and so it allows a more parsimonious framework. Taking this approach, we implicitly assume constant influence

[2]Results from a Random Walk model that includes only the seasonality components are available upon request, they are less competitive and hence are omitted here.

throughout the time scale. For example, if we model the series as an $AR(7)$ process, we implicitly assume a constant coefficient for lags 2 through 7, so the HAR model will only have 3 parameters in this case, an intercept, the first lag and the average for lags 2-7. The model is formally written as

$$\bar{y}_t = \alpha_0 + \alpha_1 \bar{y}_{t-1} + \alpha_2 \bar{y}_{t-1,\omega_1} + \alpha_3 \bar{y}_{t-1,\omega_2} + \sum_{k=1}^{K} \psi_k d_{t,k} + \varepsilon_t \tag{2}$$

where $y_{t-1,\omega_l} = \frac{y_{t-1}+\ldots+y_{t-\omega_l-1}}{\omega_l}$, $l = 1, 2$. We use $\omega_1 = 7$ and $\omega_2 = 30$, corresponding with one week and one month. $d_{t,k}$ is the exogenous variables at time t, the dummies for weekend days, dummies for month of the year and an intercept term.

3.1.2 Dynamic ARX model

An ARX(p) model for the daily spot price is defined as:

$$\bar{y}_t = \varphi_0 + \sum_{i=1}^{P} \varphi_i \bar{y}_{t-i} + \sum_{k=1}^{K} \psi_k d_{t,k} + \varepsilon_t$$

where P is the number of lags included in the model. We estimate this model with a *maximum* of $P = 14$ lags. It is dynamic in a sense that P is chosen at every point in time according to the Akaike information criterion (AIC). That is, at every time point we choose the number of lags that minimizes $AIC(p) = -2 \log L + 2k$ Where k is the number of parameters in the model and L is the likelihood value.[3] We will use this model as our benchmark for comparison.

3.2 Multivariate Models

The dataset contain 24 series and so, multivariate modelling framework requires estimation of a large number of parameters. When we consider VAR-type models we must limit the number of lags we use. The number of parameters sharply rises with the inclusion of

[3]We also performed this with the Bayesian information criterion but its results were less competitive.

an extra lag, and with it, model complexity. We use the first, second and the 7th lag. The choice is motivated by the mean reversion property observed in electricity prices (Knittel and Roberts, 2005) as well as the weekly seasonality pattern (see bottom left of figure one). Same lag structure was also used by Weron and Misiorek, (2008). .

An unconstrained model with a large number of parameters might be ill behaved when it comes to forecasting. The appended flexibility potentially creates in-sample over-fitting at the expense of forecast accuracy. Our framework is especially prone to this problem as the spot price series is the average 24 individual components, as oppose for example, to a yearly figure that is the average of four quarters.

3.2.1 VAR Type Models

Autoregressive models are a natural choice for electricity price forecasting, as they exhibit strong mean reversion properties. Consider a panel $\{y_{j,t}\}$, $t = 1, ..., T$, and $j = 1, ..., M$. Denote P and K as the number of lags and number of exogenous variables respectively. M is 24 in our case. Define $Y_t = (y_{1,t}, ..., y_{M,t})'$, let $\boldsymbol{\iota}_{M\times 1} = (1, 1, ..., 1)'$, and $D_t = (d_{1,t}, ..., d_{K,t})$ with $d_{k,t}$, a dummy variable, for example, $d_{Sunday,t} = 1$ if day t is a Sunday and otherwise $d_{Sunday,t} = 0$. Finally $X_t = (Y_{t-1}, Y_{t-2}, Y_{t-7}, \boldsymbol{\iota}, D_t)'$. We can now write the VAR model as

$$Y_t = \Phi X_t + e_t, \qquad e_t \sim i.i.N(0, \Sigma) \tag{3}$$

It is convenient to rewrite the model in even more compact form:

$$\boldsymbol{Y} = \boldsymbol{X}\Phi + \boldsymbol{E} \tag{4}$$

where $\boldsymbol{Y} = (Y_{1+P}, ..., Y_T)'$ is a $(T-P)\times M$ matrix of dependent variables, in our case the hourly price series, $\boldsymbol{X} = (X_{1+P}, ..., X_T)'$ is the $(T-P)\times(MP+K)$ matrix of explanatory variables, Φ is a $(MP+K)\times M$ coefficient matrix, and lastly $\boldsymbol{E} = (e_{1+P}, ..., e_T)'$ is the $(T-P)\times M$ error matrix.

In order to examine the contribution of using a multivariate framework we consider three VAR-type models. A model where we restrict the cross lags coefficients to zero, we dub it Diagonal VAR (DVAR). The model essentially boils down to 24 simple autoregressive models. We will see that despite the flexibility of this model where every hour is modelled separately, it does not significantly outperform the dynamic univariate $ARX(p)$ benchmark model. A model with no restriction on the cross lag coefficients, an unrestricted VAR (UVAR) and a Baysian VAR, (BVAR) which will be discussed in the following subsection. The UVAR model has many parameters, with three lags, and 14 exogenous variables, for each series we have 2064 coefficients so the model is likely to over fit the in-sample period. We use the BVAR as a simple way to shrink the estimates for the lags towards a random walk, and by that prevent the inherent over-fitting problem.

3.2.2 Bayesian VAR

Define $\boldsymbol{\alpha}_{M(MP+K)} = vec(\Phi)$, and $\boldsymbol{y}_{M(T-P)} = vec(\boldsymbol{Y})$, where $vec(\cdot)$ is the stacking operator. We can now rewrite the model as:

$$\boldsymbol{y} = (\boldsymbol{I}_M \otimes \boldsymbol{X})\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{5}$$

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

In the classical approach, this model can be estimated using standard least squares. Despite its simplicity, the large number of parameters produces high estimation uncertainty and may over-fit the data. In turn, this will affect the model forecasts. In order to deal with this caveat, we can do three things. The first is to impose some restrictions on the coefficients. The second is to reduce the dimension of the problem, and lastly, a combination of the two. The BVAR model belongs to the first category. In later subsections we present few models which belong to the other two.

We choose a Minnesota prior distribution with mean and variance so that the final estimates are shrunk, effectively mitigating the over-fitting problem. The Minnesota prior assumes that $\boldsymbol{\alpha}$ is normally distributed with $\boldsymbol{\alpha}^{prior}$ and $\boldsymbol{V}^{prior}$, the prior mean and covariance matrix for the parameters joint distribution. For $\boldsymbol{\alpha}^{prior}$, we take a value of zero for own lags of order larger than one and cross lags, thus ensuring shrinkage. A value of 1 for own first order lag coefficients. So:

$$\Phi^{ij}_{(MP+K)\times M} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

These are traditional choices for the prior means, e.g. Koop and Korobolis (2010)).[4]

The prior covariance matrix $\boldsymbol{V}^{prior}_{MP+K}$ determines the amount of tightness around the prior mean. The higher the prior variance, the closer is the final estimate to its unrestricted VAR estimate. For example, we do not impose shrinkage on the coefficients of the exogenous variables by assigning large values on their prior variances, letting the data determine their final values. So their mean posterior values are the OLS estimates. It is natural to assume smaller variance for higher order lags, reflecting the assumption that these should have smaller overall impact in prediction. The Minnesota prior assumes the prior covariance matrix to be diagonal. Let $\boldsymbol{V}_j$ be the block associated with the coefficients in equation j, and let $\boldsymbol{V}_{j,ii}$ be its diagonal elements, $i = 1, ..., MP + K$. We specify the prior variance of the coefficient associated with lag p for variable j as:

$$\boldsymbol{V}_{j,ii} = \begin{cases} \frac{\lambda_1}{l^2} & \text{for coefficients on own lag for lag } l = 1, ..., P \\ \frac{\lambda_2}{l^2}\frac{\sigma_{ii}}{\sigma_{jj}} & \text{for coefficients on cross lags for lag } l = 1, ..., P \\ \lambda_3 \sigma_{jj} & \text{for coefficients on exogenous variables} \end{cases} \tag{7}$$

We estimate σ_{jj} recursively at every time point t using the standard error of the residuals from a univariate autoregressive model for each of the 24 series. The ratio $\frac{\sigma_{ii}}{\sigma_{jj}}$ accounts for the different variability of the series. A more volatile hour will be assigned a lower

[4]We can also rely on the fact that energy prices are mean reverting, and use a value smaller than 1. We do not follow this path to avoid a somewhat arbitrary choice.

prior variance, effectively keeping coefficients of cross lags shrinkage constant across the different hours. The $\lambda' s$ are hyper parameters, they control for amount of shrinkage applied. The exact choice for these values depends on the application at hand. In this paper we simply follow **?** and set λ_1 and λ_2 to 0.5 and λ_3 to 100, which essentially is large enough not to shrink exogenous variables at all.

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

$$\boldsymbol{\alpha} \mid \boldsymbol{y} \sim N(\boldsymbol{\alpha}^{post}, \boldsymbol{V}^{post}) \tag{8}$$

with

$$\boldsymbol{V}^{post} = \{(\boldsymbol{V}^{prior})^{-1} + \widehat{\Sigma}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\}^{-1} \tag{9}$$

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

It is easy to see why the Minnesota prior is a popular choice. First, the posterior and predictive results are analytically solvable, which greatly facilitates our recursive computation. Second, there are many adjustments one can apply in practice. The choice of prior mean vector, the choice of hyper parameters and even the choice of the shrinkage structure. We use the exponentially declining weights as in the original prior, but one can address the exponent as an extra hyper-parameter and optimize its value over the in-sample period, e.g. **?** use linearly declining weights and **?** grid search over different combinations of hyper parameters. We tested other methods for choosing the λ's such as optimization and grid search. We found no evidence for a better way to chose these values so we simply follow "Occam's razor" principle here.

3.2.3 Factor Models (FM)

Another common way to account for the curse of dimensionality is through dimension reduction techniques such as Factor Models advocated, among others, by Stock and Watson

(2002). The idea is to summarise variability in the data using few G, say, linear combinations of the variables. Stock and Watson (2002) established the theoretical results for a two-step procedure where in the first step we extract the time series of the G factors $\{\widehat{F}_t\}_{i=1}^{T}$ and forecast the dependent variables in the second step. Forecasting can be done in two ways. Firstly, we can project each $y_{j,(t+1)}$ onto the space spanned by $\{\widehat{F}_t\}$ using the standard OLS, i.e. $\widehat{y}_{j,t+1} = \sum_{i=g}^{G} \widehat{\alpha}_{g,j} \widehat{f}_{t,g}$, where $\widehat{\alpha}_{g,j}$ is the OLS estimate for variable j of the marginal effect of factor g. Secondly, we can model the G factors as a VAR time series model, obtain forecasts $\{\widehat{F}_{t+1}\}$ and use these in the regression $\widehat{y}_{j,t+1} = \sum_{i=g}^{G} \widehat{\alpha}_{g,j} \widehat{f}_{t+1,g}$. The approach is also referred to as a VAR-PCA model. We performed both, we report the latter since results are slightly better. For compatibility, the factor VAR process has the same lag structure as in the other models, namely the first, second, and one week lag. The factors are extracted using the deseasonalized price series, and the forecasts are adjusted accordingly. More formally, the forecast is given by:

$$\widehat{F}_{t+1} = \delta_1 \widehat{F}_t + \delta_2 \widehat{F}_{t-1} + \delta_3 \widehat{F}_{t-6} \tag{11}$$

$$\widehat{Y}_{t+1} = \Theta \widehat{F}_{t+1} + \Gamma D_{t+1} \tag{12}$$

where in (**??**), $\widehat{F}$ on the right hand side is the estimated vector of G of principal components obtained using singular value decomposition. $\delta_{(.)}$ is a $G \times G$ estimated coefficient matrix, and as before, $D_t = (d_{1,t}, ..., d_{K,t})$ is the vector of exogenous variables. Θ and Γ are of dimensions $M \times G$ and $M \times K$ respectively. All coefficients together with the extraction of the principal components are estimated using the five year rolling window up to time t.

3.2.4 Reduced Rank Regression (RRR)

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 does just that. It has long been utilized for time series analysis (Velu and Reinsel, 1998), Carriero et al. (2011) prove consistency and provide rate of convergence for the estimates when the number of explanatory variables in the system tends to infinity. They apply the method on a large dataset to forecast economic variables. The basic idea is to impose a rank restriction on $\Phi_{(MP+K)\times M)}$, the matrix of coefficients in (**??**), and by that focus on a smaller number of underlying components. A solution can be employed using the Eckart-Young theorem. Say $\widehat{\boldsymbol{Y}}_{(T-P)\times M}$ is the matrix of fitted values given by standard OLS solution that minimizes the error matrix in (**??**), and let $U\Lambda V'$ be its singular value decomposition, where Λ_M is a diagonal matrix with the singular values arranged in decreasing sequence $\lambda_1 \geq ... \geq \lambda_M$ on its diagonal. We can now cast $\widehat{\boldsymbol{Y}}$ onto a subspace $\widehat{\boldsymbol{Y}}^s = U\Lambda_{(s<M)}V'$,where $\Lambda_{(s<M)}$ equals Λ with last $M-s$ elements on the diagonal set to zero. We can see that in contrast to Principal Component Analysis, RRR pays more attention to the output matrix $\boldsymbol{Y}$ then to the input matrix $\boldsymbol{X}$. Define $\boldsymbol{\Upsilon}_s = \sum_{i=1}^{s} \nu_i \nu_i'$, where ν_i is the ith right singular vector from the singular value decomposition of $\widehat{\boldsymbol{Y}}$, we can easily proceed to get the constrained coefficient matrix and the new fitted values through:

$$\widehat{\Phi}^{(s)} = \widehat{\Phi}\boldsymbol{\Upsilon}_s \tag{13}$$

$$\widehat{Y}_t^{(s)} = \widehat{\Phi}^{(s)} X_t \qquad t = p+1, ..., T \tag{14}$$

For more details on this procedure see **?**.

3.2.5 Reduced Rank Bayesian VAR (RRP)

So far we have outlined few models that try to avoid the over-fitting problem via shrinkage of the parameters (BVAR) or dimension reduction (RRR and FM). A model that combines these two approaches is the Reduced Rank Baysian VAR. Carriero et al. (2011) suggested a new method that implement both rank reduction and shrinkage. A Reduced Rank Bayesian VAR, or Reduced Rank Posterior (RRP). We apply 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 ν_i from the singular value decomposition of $\widehat{\boldsymbol{Y}}$ that was obtained via the UVAR model, we now use $\widehat{\boldsymbol{Y}}$ which was obtained using the BVAR model, and so our RRP estimator is

$$\widehat{\boldsymbol{\Phi}}_s^{RRP} = \widehat{\boldsymbol{\Phi}}^{BVAR}\boldsymbol{\Upsilon}_s \tag{15}$$

where here $\boldsymbol{\Upsilon}_s = \sum_{i=1}^{s} \nu_i \nu_i'$, ν_i now is the ith 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 Forecasts Combination

Forecasting performance of different approaches may vary both over time and across the different time series. 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 for example Timmermann (2005). As we shall see, it is so in our case. We report the performance of two possible ways for forecast averaging. The first is the simple average (AVE), i.e.

$$\widehat{\overline{y}}_t = \frac{\sum_{w=1}^{W} \widehat{\overline{y}}_{t,w}}{W}, \tag{16}$$

where W is the number of models used, and $\widehat{\overline{y}}_{t,w} = \frac{\sum_{j=1}^{M} \widehat{y}_{t,j,w}}{M}$, the average across the 24 hours for model w, or simply the spot forecast from model 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 and so we do not implement this model here.

Another common way to combine the forecasts is a simple linear regression, i.e.:

$$\widehat{\widehat{y}}_t = a_0 + \sum_{i=1}^{W} a_i \widehat{\widehat{y}}_{t,w}, \tag{17}$$

We estimate the coefficients using the most recent year, (365 data points). This approach has some drawbacks. Firstly, we lose the interpretation of the coefficients, as the weights can result in any value. Negative values or large absolute values are hard to interpret in this context. Secondly, there is a multicollinearity issue since it is reasonable to assume that the forecasts will be highly correlated and hence the coefficients may vary drastically with extreme positive or negative values. We therefore choose to pursue a more stable approach. We use a constrained least squares (CLS), which we can interpret with less difficulties. At every time point t we numerically solve:

$$\widehat{\widehat{y}}_t = \textstyle\sum_{w=1}^{W} a_w \widehat{\widehat{y}}_{t,w} \tag{18}$$

$s.t$

$$\textstyle\sum_{w=1}^{W} a_w = 1 \tag{19}$$

$$a_w \geq 0, \qquad \forall w \in W \tag{20}$$

We use a rolling window of one year. The first 365 days of the out-of-sample period are used to estimate the initial weights. Hence, when we compare the performance of this method with the other models, we omit the first 365 values so that all methods have the same number of forecasts for evaluation.

We also test another common weighting scheme, the inverse of the mean squared forecast error, i.e. $a_w = \frac{1/MSFE_w}{\sum_{w=1}^{W} 1/MSFE_w}$. Results from both the simple linear regression and the inverse of the mean squared forecast error are similar and therefore omitted. We prefer the CLS scheme over the simple linear regression since we can easily examine the contribution of each model. The weights are all positive and sum up to one which facilitates interpretation of model importance in the overall weighting.

4 Forecasting Results

In this section we set up the evaluation criteria and examine the forecasting results. We show there is valuable information in the intra-day hourly profile that is important for forecasting. We compare the models using the Giacomini-White test for unconditional predictive ability, and investigate their performance over time.

4.1 Forecasting performance evaluation

We examine the forecasting performance for the spot price. That is, after we obtain the forecasts for the 24 daily hours, we average them and compare with the actual spot price. In addition, it is interesting look at the performance of each model with regards to the individual hours. It is sensible to assume that the best model for the individual hours will perform best for the spot price itself. That is, if model A performs better than model B for each individual hourly price series, model A will also perform better than model B forecasting the spot price. Yet it must not be the case, especially if we do not account for the variance of each individual series. A model may perform well for volatile hours and fail for other less volatile hours such that the average across hours is a poor forecast for the spot price. We now outline few measures that attend to these issues.

The most common metrics for forecast evaluation are the root mean squared error (RMSE) and mean absolute error (MAE). In addition to these we add two more measures. The first is the mean percent error (MPE), which reflects that the absolute distance of a forecast from its observed value has different economic impact when the price level is very different. For example, a 5 NOK error when the price is 200 NOK is not the same as an error of 5 when the price is 800 NOK, the volatile nature of electricity prices highlight the need for this measure. The second is a Weighted Root Mean Squared Error (WRMSE)

,see for example Christoffersen and Diebold, 1998.

The measures are calculated as:

$$RMSE = \sqrt{\frac{1}{(T-h)} \sum_{t=h+1}^{T} (\widehat{\overline{y}}_t - y_t)^2}, \tag{21}$$

$$MAE = \frac{1}{(T-h)} \sum_{t=h+1}^{T} |\widehat{\overline{y}}_t - y_t|, \tag{22}$$

$$MPE = \frac{1}{(T-h)} \sum_{t=h+1}^{T} \frac{|\widehat{\overline{y}}_t - y_t|}{|y_t|}, \tag{23}$$

where $\widehat{y}_{t,j},\ j = 1, ..., 24$ is the hourly price forecast, $\widehat{\overline{y}}_t = \frac{\sum_{j=1}^{M} \widehat{y}_{t,j}}{M}$ is the spot forecast, T is the total number of observations, and h is the window length. When we compare the performance for the individual, we report the average across the 24 hours. We examine the stability over time in a later subsection.

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 weigh the series according to their volatility, so that the more volatile hours will not dominate the evaluation. A $WRMSE$ is calculated as $RMSE'Q$ where $RMSE_{1\times24}$ is a vector of the $RMSE$ measure given above for the individual hours, and Q is a (1×24) vector with $(\frac{var(y_j)}{\sum_{j=1}^{24} var(y_j)})^{-1}$ at its jth entry. The $WRMSE$ measure is only relevant when we look at forecasts of prices for individual hours.

4.2 Results

We start with the results for the spot price forecasts. The forecasting performance for the spot price itself is shown in table **??**. The first line is the performance measures for our ARX(p) benchmark model. The performance of the other models are presented with respect to the this model in the form of $\frac{\text{perfromance measure for specific model}}{\text{performance measure of the ARX(p) model}}$, so for example, the first 1.01 for the $RRR(1)$ model in the $RMSE$ column means that the $RMSE$ for the $RRR(1)$ model was slightly higher than 23.41.

Table **??** presents results for the forecasts of the individual hours. Every column corresponds with one of the aforementioned evaluation measures. Apart from the multivariate $WRMSE$ measure, the others quantities are calculated for each hour and averaged across the hours. The $WRMSE$ is just a weighted average of the $RMSE$ as explained in the previous section. We can see that in both tables, **??** and **??**, the Forecast Combination method performs best. On average, it provides above 15% improvement over what is achievable using a univariate framework. Among the models the $BVAR$ and the five factors model stand out as the best individual performs, but are still trailing forecast averaging.

It is interesting to examine which models performs better in which period. In order to do so, we plot the weights for each model formed via the (CLS), as the weights are determined using the accuracy of each model during the most recent 365 days. Figure **??** outlines the weights given for selected models over time. For convenience, the spot price process is plotted as well. We can see that most of the time the $BVAR$ model dominates the $UVAR$ model with the highest weight in most of the forecasting period. The RRP model, despite its mediocre individual performance, has an overall noteworthy weight in combination with the other models. We can also see that the weight of the more flexible $BVAR$ model is increased during more volatile periods of the spot price, not necessarily with the price levels, e.g. around 2004, This suggests that the information in the hourly profile is more important during these periods.

4.3 Testing For equal predictive accuracy

We now address the question of whether the difference in forecasting performance between the models is indeed significant. We use the Giacomini White test for unconditional predictive ability. The computation of the test statistic is identical the test for predictive

MODEL	*RMSE*	*MAE*	*MPE*
ARX(p)	23.41	11.93	0.054
HAR	1.33	1.44	1.432
RRR(1)	1.01	0.89	0.906
RRR(2)	0.98	0.88	0.896
RRR(5)	0.98	0.85	0.896
FM(1)	0.96	0.93	0.927
FM(2)	0.91	0.91	0.905
FM(5)	0.90	0.88	0.891
DVAR	1.007	1.03	1.026
UVAR	0.978	0.88	0.896
BVAR	0.89	0.83	0.841
RRP(1)	0.92	0.91	0.962
RRP(2)	0.91	0.90	0.954
RRP(5)	0.91	0.90	0.955
AVE	0.88	0.82	0.834
CLS	**0.84**	**0.80**	**0.819**

Table 2: Performance comparison between the different models, for the spot price. The first line is the benchmark ARX(p) model, the rest of the models are presented as a ratio: $\frac{\text{perfromance measure for specific model}}{\text{performance measure of the ARX(p) model}}$. RRR: Reduced Rank Regression with rank in parentheses, FM is factor model with the number of factors in parentheses, DVAR: Diagonal VAR, UVAR: Unrestricted VAR, BVAR: Bayesian VAR, RRP: Reduced Rank Posterior with rank in parentheses, HAR: Heterogeneous Autoregressive model, 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 first estimation period starts at 4th may, 1992.

MODEL	*WRMSE*	*RMSE*	*MAE*	*MPE*
RRR(1)	33.08	34.56	16.07	0.080
RRR(2)	28.82	31.07	13.83	0.069
RRR(5)	27.44	29.81	12.73	0.062
FM(1)	30.67	32.23	15.31	0.074
FM(2)	27.69	29.52	13.67	0.065
FM(5)	26.80	28.73	12.94	0.062
DVAR	28.45	30.62	14.03	0.067
UVAR	27.17	29.56	12.47	0.061
BVAR	26.36	28.59	12.30	0.058
RRP(1)	31.60	33.11	16.15	0.082
RRP(2)	28.33	30.46	14.70	0.072
RRP(5)	26.94	29.18	13.73	0.066
AVE	27.83	29.83	13.16	0.065
CLS	**25.29**	**27.31**	**11.97**	**0.057**

Table 3: Performance comparison between the different models. The quantities were calculated for each individual hour and were averaged. RRR: Reduced Rank Regression with rank in parentheses, FM is factor model with the number of factors in parentheses, DVAR: Diagonal VAR, UVAR: Unrestricted VAR, BVAR: Bayesian VAR, RRP: Reduced Rank Posterior with rank in parentheses, HAR: Heterogeneous Autoregressive model, 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 first estimation period starts at 4th may, 1992.

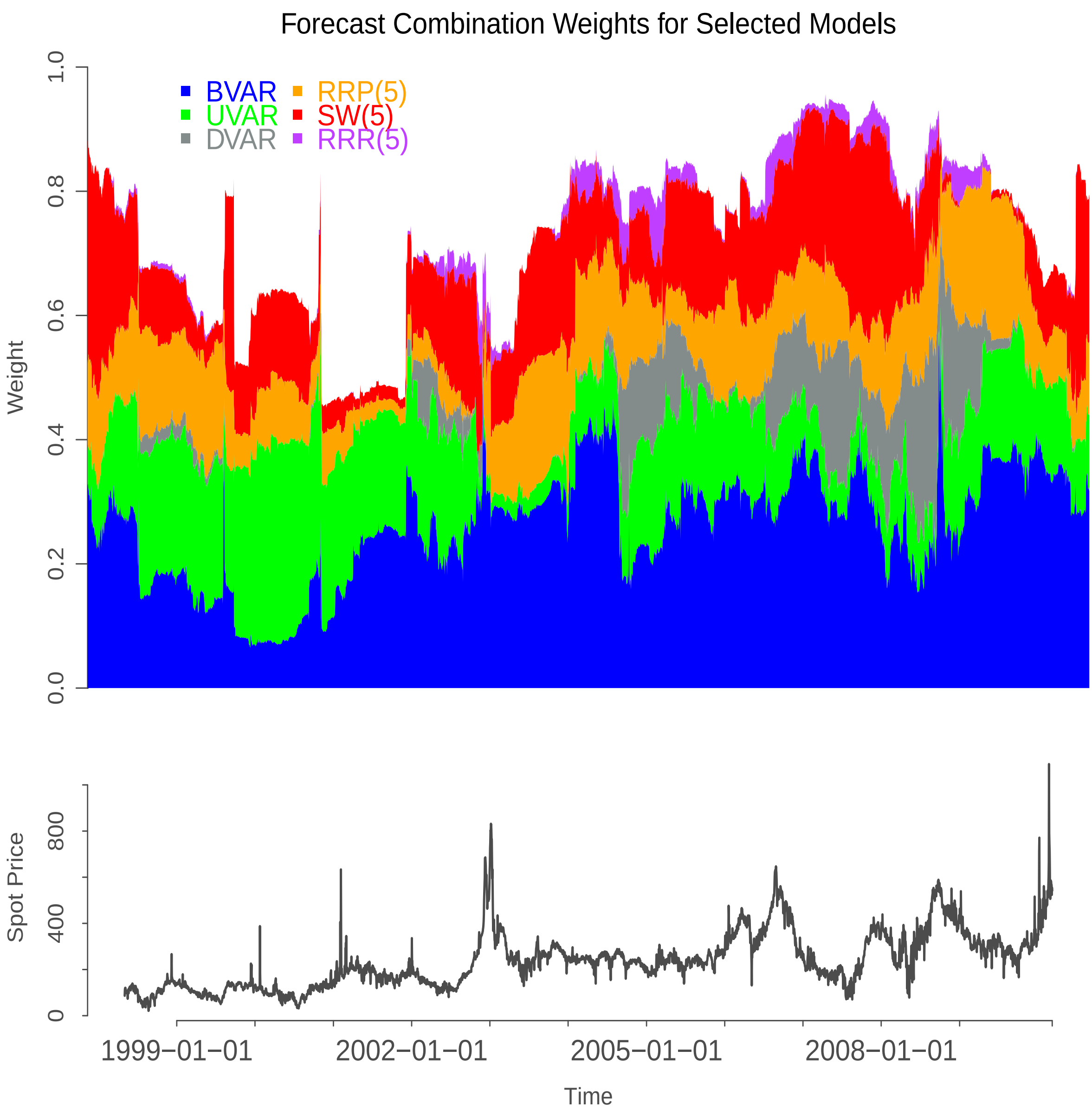

Figure 3: weights given to each model at each time point. The weights are determined using a constrained least squares procedure over the most recent 365 days in a rolling window scheme, the constraints are: weights are positive and sum up to one (not to clutter the graph, not all model weights are presented). The bottom panel is the spot price itself over time.

accuracy in Diebold-Mariano (1995). However, Giacomini and White (2006) generalize the test and develop the theoretical basis for a comparison between *methods*, as opposed to *models*. As a results we can compare between nested and non-nested models, and allow for parameter estimation uncertainty. in the forecast evaluation the test statistic is computed as

$$\delta_t = L(e^1_{t+1|t}) - L(e^2_{t+1|t}) \tag{24}$$

$e^w_{t+1|t}$ is the forecast error from method $w = 1, 2$, at time point t. For a loss function $L(\cdot)$, the null hypothesis is $E[\delta_t] = 0$. Due to the spikes in the data we prefer to use the absolute loss function which is more robust to outliers. Table **??** presents a pairwise comparison between the different methods for the forecasts of the spot price. The statistic was computed such that for entry ij, $\delta_t = L(e^i_{t+1|t}) - L(e^j_{t+1|t})$, so a positive statistic at the entry ij means that the absolute residual from model i is on average, larger than the absolute residual from model j.

We can observe few things. First, at reasonable confidence levels, the $BVAR$ model significantly outperforms its unrestricted version, the $UVAR$ model, which demonstrates the efficiency of the shrinkage procedure and appropriateness in this case, it also significantly outperforms the benchmark model. The five factor model $FM(5)$ is also significantly better than the benchmark. The forecast combination method with the weights determined by the constrained least squares has the largest absolute statistic value out of those compared with the benchmark model for the absolute loss function and is significantly different from the $ARX(p)$ model at 99% confidence levels.

4.4 Stability Analysis

In previous sections we have shown that there is significant advantage to be gained by implementing a multivariate framework. We now address the robustness of this result. Figure **??** presents a rolling ratio between selected models and the benchmark model. We use the MPE measure and a three year window length. We see that we can gain around

	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	
ARX(p)	6.22	4.09	-1.93	3.18	5.58	4.08	6.96

Table 4: Giacomini White test statistics for the absolute loss function. The critical value for the 95% confidence level is 1.65 or 1.96 for a two sided test. The statistic was computed such that a negative statistic at the entry ij means that the absolute residual from model i is on average, smaller than the squared residual from model j

15% improvement with respect to the benchmark, and that this gain is not attributed to a specific sample period but is relatively stable. The Diagonal VAR, being restricted from using the additional source of information, namely the cross sectional relation, performs slightly worse then the benchmark. The FM(5) model was performing quite well up prior to 2003 yet in recent decade, its performance deteriorated with respect to the benchmark. This fact which displays its average good performance in Table **??** in a slightly different light.

The $UVAR$ manages to outperform the factor model most of the time. We can further improve upon that using the $BVAR$ model which mitigates the over-fitting problem. The Forecast Combination using Constrained Least Squares method is dominant throughout the sample. The improvement brought about by this methods is up to 20% at times and never under-performs the benchmark.

5 Conclusion

The results presented in this paper suggests that, for forecasting purposes, it is beneficial to exploit the information embedded in the cross correlation of the hourly price series. This can be done by moving towards a multidimensional framework in which we forecast

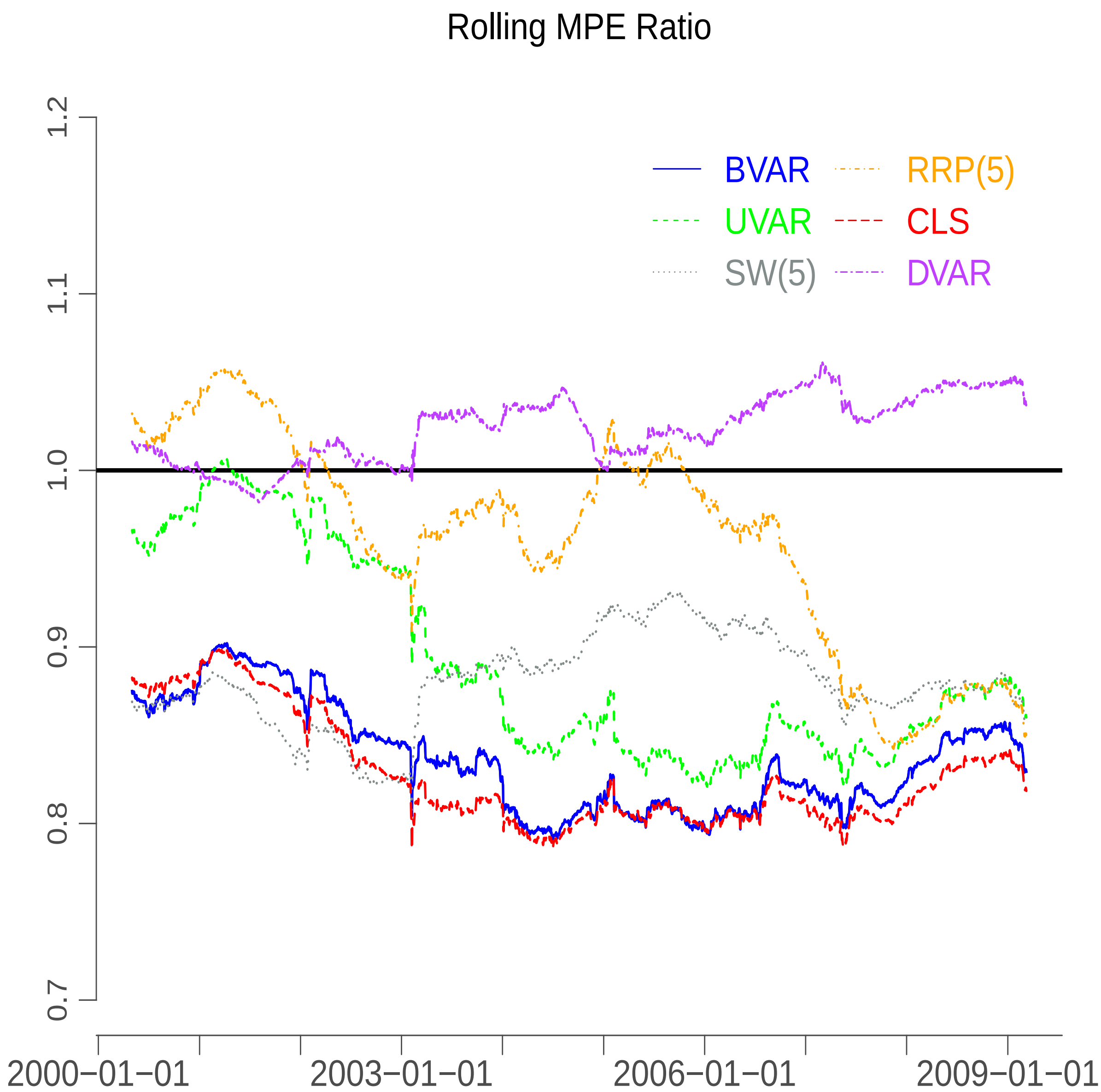

Figure 4: Three years rolling MPE measure. The ratio between selected models and the $ARX(p)$ benchmark. The horizontal line at value of one represents equal MPE between the specific model and the benchmark.

each individual hour separately while allowing cross relation between the hourly prices. Spot price forecast is obtained by taking the average across the 24 hourly prices. Diagonal VAR model does not outperform a univariate benchmark, however, allowing for cross lags effects improves performance while using dimension reduction techniques and forecasting averaging produce significant improvement of about 15% accuracy compared with a flexible univariate approach.