Forecasting Hourly Electricity Prices using Principal Components

Kimberley G. Reeder

Erasmus University Rotterdam

June, 2012

Abstract

This research considers forecasting 1 day-ahead hourly electricity prices in the Norwegian power market, after a principal component analysis has been applied on the data. As a way of gaining forecast accuracy, combinations of weighted individual forecast were also considered. Comparisons between the accuracy of five distinct models were made. A base model; in which principal component analysis was applied to the dataset comprising all 24 hourly prices, was used as standard to compare with the alternatives, by means of three evaluation criteria. Ultimately, the results were somewhat the same for the peak/off-peak hour model and the 4 intra-day period model: both were outperformed by the base model, mainly for periods of high volatility. The weighted principal components model simply generated poor results. The combination of weighted base model with the weighted peak/off-peak model was the only model which generated better for volatile periods, but this at the expense of losing accuracy in other periods.

Table of Contents

1. Introduction 1

2. The Data 3

3. Methodology 5

3.1 Principal Component Analysis 5

3.2 Forecasting 8

3.2.1 The base model 8

3.2.2 The peak/off-peak hour model 9

3.2.3 The 4 intra-day periods model 10

3.2.4 Combining individual forecasts by assigning weights 10

3.2.5 The combined weighted base and weighted peak/off-peak model 12

3.3 Forecast evaluation criteria 13

3.4 Comparing predictive accuracies 14

4. Results 15

4.1 Optimal number of principal components 15

4.2 Forecasts peak/off-peak hour model 16

4.3 Forecasts 4 intra-day periods model 18

4.4 Forecasts combined weighted base and weighted peak/off-peak model 20

4.5 Forecasts weighted principal components 21

5. Conclusion 23

6. References 25

7. Appendix 27

Appendix A 27

Appendix B 28

Appendix C 30

Appendix D 31

Appendix E... 33

1. Introduction

Electricity is together with oil and gas the most frequently used energy source that our civilization depends on. It is a vital product for the economy we live in. Millions of industrial and residential end-users rest on it at every particular instant of the day. Back in the 1990's there was a start of the global deregulation process in the electricity industry. Ever since, electricity prices were rapidly founded on the market rules of supply and demand. New companies entered the industry and exchange, where applicants could trade in deals that met their needs were made. The energy market that traditionally comprised of a few producers and a dozen of distributors would be different from how it was ever recognized. In order to exploit the broader variety of choices, lowest prices and the best services provided by the suppliers, end-users have to turn to action. They have to make buying decisions in line with their cost and consumption preferences.

Electricity prices are described by a unique feature that cannot be perceived in any other market. Extreme price movements take place commonly and stem from the non-storability of electricity. Given the lack of storage capability in electricity markets, it is impossible that inventories can function as a buffer between supply and demand discrepancies. As a consequence, it is essential that there is a perfect sense of balance between the amount of electricity that is produced by the power manufacturers and injected on the network transmission (supply-side), and the amount of power drained from the power centrals by distributors who sell the electricity to the end-users (demand-side). These parties need to predict the demand and/or supply as accurate as possible to secure the balance. Demand forecasting is an important facet in the development of any model for electricity planning. The form of the demand is contingent on the kind of scheduling and the accuracy that is required. The emphasis varies from minutes to several hours in the future. The predictions are essential as inputs to scheduling processes for the generation and transmission of electricity. The predictions help in defining which devices to activate in a particular period, to minimize costs (Taylor, Menezes, McSharry, 2006).

The aim in this research is on forecasting 1 day-ahead hourly electricity prices as accurate as possible. The answer on how many explanatory variables to include in the estimation model is a consideration between bias and precision. Apart from that, parsimony also plays a role meaning that in general we prefer small models with a limited number of explanatory variables. This is an important issue throughout the research, as the dataset which will be worked with consists of prices on each hour of the day, which can be considered redundant, as 24 daily hours as explanatory variables may cause over fitting when forecasting out-of-sample prices for each hour. Reducing the amount of explanatory variables by means of Principal Component Analysis (PCA hereafter) may yield better results on forecasting.

Furthermore, in their research Bates & Granger (1969) have proven that forecast accuracy can be improved by combining individual forecasts. Taking advantage of their insight, this research will adopt techniques of combining forecast outcomes, with the intention of gaining forecast accuracy.

Summarized, the research question is whether better accurate forecasts of the hourly electricity spot prices can be achieved after dimensionality reduction and by combining individual forecasts.

2. The Data

The dataset which has been examined finds it source at Nord Pool Spot. Nord Pool Spot runs the leading market in Europe for buying and selling power. They offer day-ahead and intraday markets to their customers, but also operate the N2EX in the UK market together with NASDAQ OMX. There are 370 companies from 20 different countries active on the market. In 2011 alone, 316 Terawatt Hours (TWh) were traded at Nord Pool Spot, which is equal to the power consumption of Oslo for 40 years. Nord Pool Spot is completely owned by Nordic transmission system operators, each of them having their share. Members of Nord Pool spot are mainly power producers, suppliers and traders. But also large end-users, who trade on the markets and buy directly from Nord Pool Spot instead, are acquainted as members. On the website of Nord Pool spot (http://www.nordpoolspot.com/), more information can be found such as the market overview.

The dataset which has been analyzed, comprise of prices for 24 hourly time series, one for each hour during a day, in Norwegian kroner (Kr) per MWh, for the period starting August, 12 2009 and ending March, 9 2010, totaling 210 day-ahead prices. In the day-ahead market an auction of power for delivery on the following day is set. The prices in the day-ahead market are calculated according to supply, demand and transmission capacity (how much power can be transported from one place to another). This is Europe's most liquid day-ahead market, producing an accurate and reliable reference price. The dynamics of the electricity prices for two hours with clearly distinct features are plotted in figure 1, for the 210 days. Some statistical properties of the data are presented in Appendix C- table C1. It can be observed that the means of the hourly electricity spot prices during the sample period varies between 326.09 Kr/MWh for the 18th hour and 240.56 Kr/MWh for the 4th hour. Indicating that power demand during these hours is relatively high and low, respectively. The hourly spot prices reached a maximum value of 1042.32 Kr/MWh for the 18th hour and 313.84 Kr/MWh for the 4th hour respectively. The standard deviation for the 18th hour is very high (97.70 Kr/MWh), indicating that prices during this hour are very volatile. As the values for the skewness for the 10th till 21th hour all have positive signs, this indicates that high extreme values are more likely to occur during this period of the day, than low extreme values, as opposed to the remaining hours.

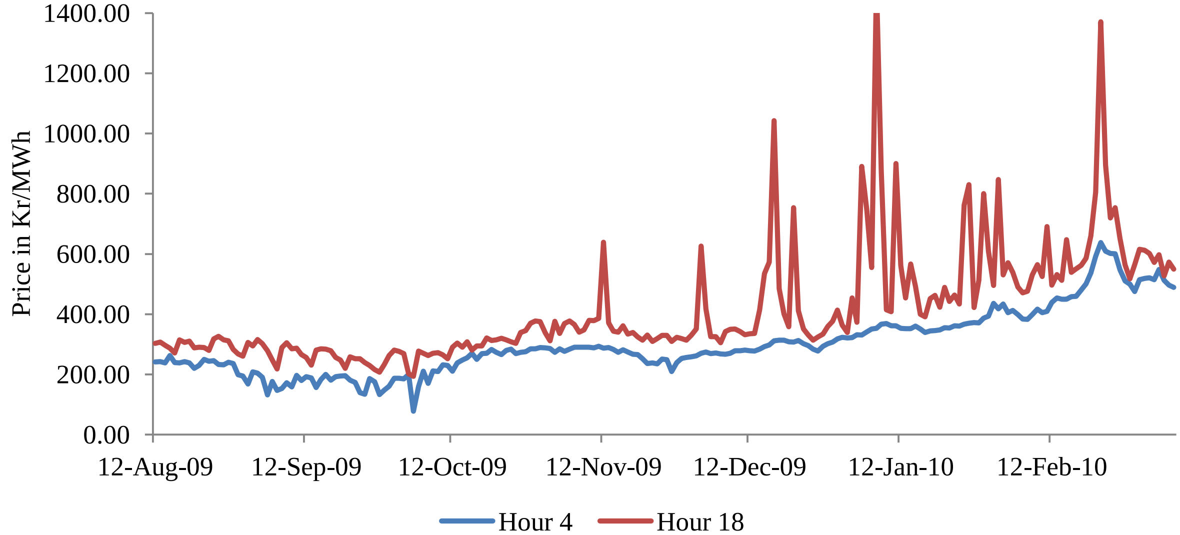

Figure 1: This graph displays the dynamics of the electricity prices in Kr/MWh for the 4th and 18th hour over the period August, 12 2009 - March, 9 2010.

3. Methodology

This section describes the methodology used in the research. First in section 3.1 PCA will be discussed in detail, followed by the forecasting methods in section 3.2, which have been applied to the research data. Section 3.3 describes the performance measure to evaluate the forecasting results. And finally in section 3.4 the Diebold-Mariano test statistic is described which tests the difference in forecast predictive accuracies after forecasts have been made with two different models.

3.1 Principal Component Analysis

PCA, first introduced by Pearson (1901) is a method of data dimension reduction without sacrificing too much information inherent to the original data. Dimensionality is here defined as the number of independent variables. Simply stated, by making use of the PCA procedure, one tends to find a set of linear combinations of factors $X_t$which are uncorrelated and explain most of the variance in X_t. Technically stated PCA is based on a procedure where the X_t observations are split into P linearly uncorrelated factors called principal components. By definition the first principal component explains most of the variance of the data; the second principal component explains the second largest part of the variance and the third component explains a smaller portion of the variance in X_t and so on. The percentage of variance explained decreases with each principal component ranking lowest relative to the previous component. In theory it is required that the amount of principal components is less than or equal to the amount of original variables in X_t. The following part describes the PCA procedure more in detail.

Suppose a dataset comprises a set of K variables for the hourly electricity prices. The purpose is to forecast the hourly electricity spot prices and each hour is treated as an independent variable with $K = 1, \dots ,24$. The emphasis is to reduce the data dimensions by describing the variation in the electricity prices in terms of g factors $g \ll K$. The fundamental thought is to describe the (co) variation in the K hourly electricity prices $X_{1t}, \ldots, X_{Kt}$ with a number of g factors, subject to $g \ll K$. In formula:

$$X_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{ig} f_{gt} + \epsilon_{it} \quad \text{eq. 1}$$

for i = 1,..,T

The ‘factors’ f_t are the linear combinations of the hourly electricity prices X_{it}, β_{ij} the respective factor loadings for the electricity price of day i on the factor j, and ϵ_{it} the idiosyncratic component. Here, both the factors and factor loadings are assumed to be unobserved. In general those factors f_t are chosen in such a way that they describe most of the (co)variation in the electricity prices of the original dataset X_{it}. These factors turn out to be the principal components of the covariance matrix of X_{it}.

Let $\hat{V}$ denote the KxK sample covariance matrix of $X_t = (X_{1t}, X_{2t}, \ldots X_{Kt})'$, where $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ represents the vector of the sample mean. Then the (co)variance matrix can be derived by

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T} (X_t - \bar{X})\,(X_t - \bar{X})' \quad \text{eq. 2}$$

An eigenvector e_i of the matrix V is such that $V e_i = \lambda_i e_i$ for some constant λ_i, called the eigenvalue belonging to e_i. If V is the covariance matrix of X_t, the matrix E (containing the eigenvalues) transforms the correlated electricity prices X_t into orthogonal variables $P_t = E' X_t$, that is

$$P_{it} = e_{i1} X_{1t} + e_{2t} X_{2t} + \cdots + e_{Kt} X_{Kt} + \epsilon_{it} \quad \text{eq. 3}$$

The covariance matrix of P_t is diagonal with λ_i being the variance of P_{it}, that is

$$\begin{aligned} Var[P_{it}] &= E[P_t P_t^{'}] \\ &= E[E' R_t R_t^{'} E] \\ &= E^{'} V E = E^{'} E \, \Lambda = \, \Lambda \end{aligned} \qquad \text{eq. 4}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \dots \lambda_K)$. The sum of the eigenvalues λ_i is equal to the total variance of X_t.

Let $(\lambda_1, e_1), \dots, (\lambda_K, e_K)$ be the eigenvalue-eigenvector pairs of $\widehat{V}$, ordered according to increasing values of λ_i, the fraction $\frac{\lambda_i}{\sum_{j=1}^{K} \lambda_j}$ is then the fraction of the total variance in X_t explained by the i^{th} principal component. If we manage to find $g \ll K$ such that the fraction $\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_g}{\lambda_1 + \lambda_2 + \cdots + \lambda_K}$ is reasonably large, in that case we obtain the following approximation

$$X_{it} \approx \beta_{i1} P_{1t} + \beta_{i2} P_{2t} + \cdots + \beta_{ig} R_{gt} \qquad \text{eq. 5}$$

The principal component P_{jt} may be interpreted as factors.

The amount of components obtained in after PCA is equal to the amount of observed variables being analyzed, meaning that an analysis of 24 hours results in 24 principal components. However, as before mentioned only the first few components account for a large amount of the variance in the original data, so only these first few components are retained and used for further analyses.

There are several sophisticated methods for automatically determining the optimum number of components (Hannes & Jurgen, 2009). The scree-test which was first proposed by Cattel (1966) is used to determine the number of principal components to include in the regression. It is a graphical method where the eigenvalues are plotted in descending order of their magnitude against their associated component number. The factorial scree can be found at the point where the plot of the eigenvalues does not decrease by a large amount.

3.2 Forecasting

This section describes the forecasting method applied, on forecasting the 1 day-ahead hourly electricity prices after data reduction. First it should be mentioned that the original dataset covered a period of 18 years running from April 1992 – September 2010. This dataset has been reduced to a subset of 210 days (30 weeks), including only the most recent available data. This choice has been made based on techniques adapted from former research by Taylor, Menezes, & McSharry, 2006. They considered 30 weeks of hourly observations for short-term forecasting of electricity demand in Rio de Janeiro, Brazil. Moreover, as this research considers a very short forecasting period of 1 day-ahead, excessive amount of historical data might be considered as irrelevant.

3.2.1 The base model

In the first forecasting method, PCA was applied on all the variables of X_t with $t = 1, \dots, 24$. As explained in the previous paragraph, principal components of X_t were extracted from the (co)variance matrix of X_t. Consecutively, a principal component regression (PCR) for each hour on the transformed dataset was made. In formula the regression model is specified as:

$$y_{t+1} = \alpha_0 + \alpha_1 P_{1t} + \dots + \alpha_g P_{gt} + \varepsilon_t \quad \text{eq. 6}$$

Subsequently, 1 day-ahead static forecasts $\hat{y}_{t+1}$were made ,where y_t denotes the hourly electricity price on day t, starting December, 30 2009 and extending through March, 9 2010 (10 weeks) making use of a moving window. Forecasts which were obtained according to this method will be further referred to as $\hat{y}_{\text{complete}}$.

3.2.2 The peak/off-peak hour model

In search for an alternative and more accurate forecasting method, the data was evaluated regarding the intra-day cycles of the of the electricity spot prices. In practice the daily electricity demand is different for particular hours of the day. Specifically, electricity prices exhibit a nature of intra daily profile, i.e. electricity demand during the day is relatively high compared to electricity demand during the nights, respectively defined as the peak and off-peak hours. The peak hours are usually defined over the time period of 8:00 am till 8:00 pm, for business days only and the remaining hours account for the off-peak hours. Thus peak hours are represented by $X_{t,peak}$ where $t = 9, ..20$ and off-peak hours by $X_{t,off-peak}$ where $t = 21, ..., 24$ and $t = 1, ..., 8$. For convenience, no distinction was made between weekdays and weekend days.

In same line as the first forecast method, PCA was applied independently to the peak and off-peak hours, i.e. to each of the two different blocks, as each block may yield a different number of principal components, depending on the covariance matrix structure of the block. The following regressions were applied on the two blocks:

$$y_{t+1}^{peak} = \beta_0 + \beta_1 P_{1t} + ... + \beta_g P_{gt} + \kappa_t \quad \text{eq. 7}$$

$$y_{t+1}^{off-peak} = \gamma_0 + \gamma_1 P_{1t} + ... + \gamma_g P_{gt} + \vartheta_t \quad \text{eq. 8}$$

Consecutively the parameters obtained from these regressions were used to forecast the 1 day-ahead hourly electricity prices independently. The forecasts are denoted as $\hat{y}_{peak}$ and $\hat{y}_{off-peak}$ for respectively the peak and off-peak hours. Together, these two individual forecasts will be referred to further throughout the thesis as $\hat{y}_{blocks}$.

3.2.3 The 4 intra-day periods model

This method of forecasting makes use of the patterns in the in-sample price data, as was also the case in the previous paragraph. The purpose was to find patterns or group of hours in which the prices were more or less evenly distributed. After examining the pattern, the data was split into 4 periods, with which in the same line as the base and peak/off-peak hour model, forecasts were processed.

3.2.4 Combining individual forecasts by assigning weights

In the previous paragraphs forecasts of individual intra-day blocks were made, to take into account the difference in the electricity demand. Bates and Granger (1969) proposed a way of combining forecasts by assigning weights to individual forecasts. They stated that forecast combinations could improve the performance accuracy of individual forecasts. The interest is in cases in which two (or more) forecasts have been made of the same occasion. Typically, the attempt to make, is to discover which is the better (or best) forecast; the better forecast is then accepted and used, the other being thrown out. Whereas this may have certain merit, this is not a sensible practice if the objective is to create as good a forecast as possible, since the unwanted forecast almost always encloses some valuable independent information. This independent information may be of two kinds:

(i) One forecast is centered on variables or information that the other forecast has not deliberated.

(ii) The forecast makes a different assumption about the form of the association between the variables.

In following their study, an attempt has been made to evaluate combinations of forecast results arising from different forecast methods. This research has adapted the first forecasting method, by focusing on different amounts of predictor variables in each of the forecasting model. For each forecast an assumption was made about the amount of principal components, different from the amount chosen for the other forecast. For example, when making two forecasts, the first forecast was made by using 2 principal components whilst the second forecast was made by using 3 principal components. This way all the possible combinations regarding the number of principal components could be tested. However there is no point in making a forecast by for example combining forecasts made by respectively 24 and 23 principal components. This would not be a sensible reduction of the amount of variables in the original data. For this reason a benchmark regarding the maximum allowable amount of principal components had to be set. The maximum amount of principal components by which the forecasts were carried out was chosen to be 12.

In general, each of the predictions could be given an equal weight however one would desire to assign a higher weight to the set of forecasts which appears to enclose the lower errors. There are several methods of defining these weights, and the purpose was to select a method which was expected to produce low errors for the combined predictions. It was presumed that the performances of the individual predictions were stable over time and that the variance of the two forecasts errors could be signified by σ_1^2 and σ_2^2 for all values of time *t*. It was further also assumed that the two forecasts were unbiased. The joint predictions could then be found by a linear combination of the two sets of predictions, giving a weight k to the first set of predictions and a weight $(1-k)$ to the second set, accordingly making the combined predictions unbiased. The variance of errors in the combined predictions σ_c^2 could then be written by:

$$\sigma_c^2 = k^2\sigma_1^2 + (1-k)^2\sigma_2^2 + 2\rho k\sigma_1(1-k)\sigma_2 \qquad \text{eq. 9}$$

where k is the weight given to the first set of predictions and ρ is the correlation coefficient between the errors in the first set and in the second set of predictions. The choice of k was made so that the errors of the combined forecasts were small: more precisely, the overall variance σ_c^2 was chosen to be minimized. Differentiating with regard to k, and equating to zero, the minimum of σ_c^2 was obtained when:

$$k = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad \text{eq. 10}$$

It can be revealed that if k is obtained by eq. 10 the value of σ_c^2 is no greater than the smaller of the two individual variances. The combined forecast for time period T, C_T is defined by:

$$C_T = k_T f_{1,T} + (1 - k_T) f_{2,T} \quad \text{eq. 11}$$

Where $f_{1,T}$ and $f_{2,T}$ denotes the forecasts at time T for the first and the second set of forecasts, respectively. Further each forecast is carried out by the formula in Eq. 6.

3.2.5 The combined weighted base and weighted peak/off-peak model

By means of this method, the drive was to combine the forecast of the base model from section 3.2.1 with the forecasts from the peak/off-peak model from section 3.2.2. This was accomplished by assigning weights to each set of the forecasts and combining them to one forecast. This method is an extension to the method in the previous paragraph in which the calculation steps for the determination of the weights were disclosed.

As said, the combination of the two depicted models is realized by combing the forecast obtained from the complete dataset($\hat{y}_{\text{complete}}$) with the forecast obtained from the peak and off-peak hours $\hat{\mathrm{y}}_{\mathrm{blocks}} = \begin{bmatrix} \hat{\mathrm{y}}_{\mathrm{peak}} \\ \hat{\mathrm{y}}_{\mathrm{off-peak}} \end{bmatrix}$. Again the overall variance σ_{c}^2 is minimized when the weights are obtained according to equation 10. Eventually the weights were used in the combined forecast equation C_T for time period T:

$$C_T = k_T f_{complete,T} + (1 - k_T) f_{blocks,T} \quad \text{eq. 12}$$

3.3 Forecast evaluation criteria

Forecast evaluation criteria are explained and discussed in this section, because it is of great interest to compare alternative models by their forecast performance. The forecast performances of the different models were evaluated by means of 3 criteria, namely the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE.).

The RMSE is a quadratic scoring rule which measures the average magnitude of the error.

$$RMSE = \left(\frac{1}{n}\sum_{h=1}^{n}(y_{m+1} - \hat{y}_{m+1})^2\right)^{1/2} \quad \text{eq. 13}$$

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. All the individual differences are weighted equally in the average by averaging over number of observations in the hold out sample.

$$MAE = \frac{1}{n}\sum_{h=1}^{n}|y_{m+1} - \hat{y}_{m+1}| \quad \text{eq. 14}$$

The MAPE on the other hand is a measure of accuracy for constructing fitted time series values in statistics, specifically in trend estimation. It expresses accuracy as a percentage, and puts a penalty to large errors if they are produced by relative large observations.

$$MAPE = \frac{1}{n}\sum_{h=1}^{n}\left|\frac{y_{m+1} - \hat{y}_{m+1}}{y_{m+1}}\right|$$ eq. 15

3.4 Comparing predictive accuracies

Clearly outcomes with small values for RMSE, MAE and MAPE are preferable, but the statistical significance of the difference in the forecast errors between two forecast methods is equal likely preferred. This was tested by means of the Diebold-Mariano test, which comes down to calculating the loss differential between 2 models and testing whether this value differs significantly from 0.

Let $\hat{y}_{i,m+1|m}$ and $\hat{y}_{j,m+1|m}$, be 2 different 1 step-ahead forecasts from models i and j respectively and let $e_{i,m+1|m} = y_{m+1} - \hat{y}_{i,m+1|m}$ and $e_{j,m+1|m} = y_{m+1} - \hat{y}_{j,m+1|m}$ be the corresponding errors. The loss differential can then be defined as:

$$d_{m+1} = e^2_{i,m+1|m} - e^2_{j,m+1|m}$$ eq. 16

Given a set of n 1 step-ahead forecasts the Diebold-Mariano test statistic is then defined by

$$DM = \frac{\bar{d}}{\sqrt{\frac{V(\hat{d}_{t+1})}{n}}} \sim N(0, 1)$$ eq. 1811

where $\bar{d}$ denotes the sample mean of d_{m+1} and $V(\hat{d}_{m+1})$ is an estimate of the variance of d_{m+1}, which can be calculated as

$$V(\hat{d}_{m+1}) = \frac{1}{P-1}\sum_{t=m}^{T+n-1}(d_{t+1} - \bar{d})^2$$ eq. 1912

4. Results

In this chapter the results will be presented. First the optimal number of components will be discussed for differenr forms of the dataset, followed by the forecast results obtained with the five forecasting techniques, namely forecasts results regarding the base model, the peak/off-peak hour model, the 4 intra-day period model, the combined weighted base and weighted peak/off-peak hours model and the results of weighted principal components model. Each of the forecasts will be evaluated using different forecast evaluation criteria. Finally the results of the Diebold-Mariano test will be discussed. Detailed forecasting results regarding the different models can be foun in Appendix D.

4.1 Optimal number of principal components

The optimal number of components changes each time the form of $X_t$of the in-sample data changes. Regarding the PCA on the whole in-sample dataset without making any assumptions about distinct patterns between hours, the optimal number of components were derived according to the so called scree-plot. Scree plots derived from the results have indicated that the contributions to the explanation of the variability of the data, decreased drastically after the 2nd or 3rd principal component, concluding that 3 principal components provided a reasonable summary of the data. After deriving the principal components for the peak hours the so-called ‘elbow’ of the line could be noticed at the point of 2 components, indicating that 1 component was sufficient while the remaining components may be interpreted as redundant. However in this case 2 components will be considered important too, as its fraction of the total is substantial. The same holds for the results of the off-peak hours; after analyzing the scree plot and eigenvalues, again 2 components were selected, hence 1 and 2 components account for the largest part of the variance explained. The scree plots and tables containing the principal components and eigenvalues can be found in Appendix A, table A1-A3 and Appendix B, figure B1-B3.

The optimal number of component changes also within each alternative forecast model, these will be mentioned the following paragraphs.

4.2 Forecasts peak/off-peak hour model

Figure 2 displays the difference in the error evaluation criteria between the forecasts based on the model complying eq. 7 and 8 and the base model from eq. 6. It is noticeable that the forecasts differences between the two models during the peak hours are very volatile while the forecasts differences during off-peak hours are behaving more or less constantly. Right after the start of the peak hours, at 8:00 the difference forecast error measurements RMSE and MAE are increasing strongly and after 9:00 all three measurements altogether rise to an optimum at 16:00. This indicates that the predictive accuracy of the peak/off-peak hour model is considerably less accurate during these hours. From hours 16:00 on till 19:00 the difference in the predictive accuracy criteria RMSE, MAE and MAPE is decreasing, implying that the accuracy of the peak/off-peak model is getting closer to the predictive accuracy of the base forecast model (the difference between the two models is moving towards the zero level). However from 19:00 till 20:00 again the predictive accuracy of the peak/off-peak model is getting worse as the distance of the predictive accuracies between the peak/off-peak model and the base model is positive and increasing, in favor of the base model. After 20:00 the difference between the peak/off-peak and base model is decreasing quickly in predictive accuracy as the measurement errors of the peak/off-peak hour model are decreasing and the errors for the base model are increasing. The situation gets better as after 23:00 the peak/off-peak hour model is predicting better as the difference of the evaluation criteria falls below the zero, i.e. the base model is outperformed by the peak/off-peak model. The peak/off-peak model is performing best between 03:00 and 04:00. Clearly on average the peak/off-peak hour model performs best when applied to the off-peak hour period. In a sense and on intuition, the off-peak hours are less volatile in electricity demand compared to the peak hours. In the peak hours, there is a strong electricity demand. The challenge is in making as accurate as possible forecast concerning the peak hours. However as far as the results show, the current method has not offered bright outcomes and this method may be further elaborated on.

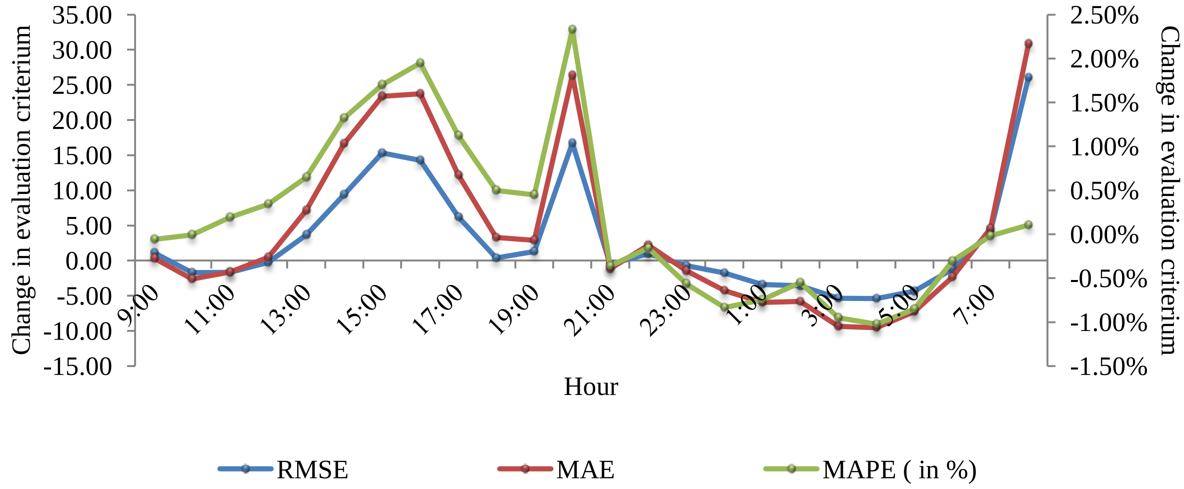

Figure 2: This graph displays the difference in the forecast error evaluation criteria between the peak and off-peak model and the base model over the period December 30, 2009- March 9, 2010.

The Diebold Mariano test statistic takes on a value of -1.06, with a corresponding p-value of 0.28 indicating that the forecasts based on the peak and off-peak hours do not generate significantly better forecasts than the base forecast when considering all the hours of the day, at a 5 percent significance level.

4.3 Forecasts 4 intra-day periods model

Figure 3 displays 4 different blocks. Each block contains absolute average values of price changes between the following and preceding hour which are more or less the same. The price change is plotted against the hours, to eliminate the trend in the data points to get a clearer picture of the association between the average prices from the in-sample data.

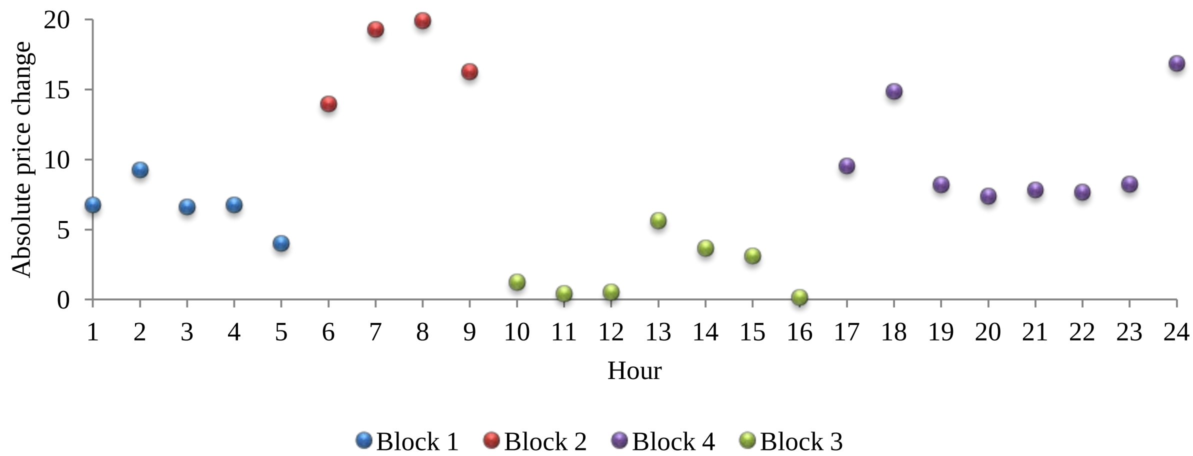

Figure 3: The figure displays the absolute average price change between hour t and hour $t-1$ over the estimation period running from August, 12 2009 – December, 29 2009.

It can be clearly seen that one can distinguish 4 different intra-day periods in which the prices are more or less evenly distributed. Starting from this empiric finding, the assumption is made that these 4 intra-day periods have to be forecasted separately and independently by means of the model in eq. 6. The forecast results based on the outcome from 4 intra-day periods are presented in figure 4. These are the differences in the error evaluation criteria between the forecasts obtained from the 4 intra-day period model and the base model. Compared to the outcomes in figure 2 of the previous analysis, the results in figure 4 are on average slightly worse for some intra-day periods. For the hours running from 05:00 till 09:00, the 4 intra-day period model seems to generate larger forecast errors than the base forecast errors, as the difference is positive and much larger than the difference depicted in figure 2 (maximum difference value of almost 70 for MAE and 5% for MAPE in figure 4). Over the period from 09:00 till 18:00 the differences are decreasing steadily and constantly, in contrast to the volatile development of the evaluation differences line in figure 2. From 18:00 till 05:00 the differences are negative in favor of the 4 intra-day periods model, which generated smaller forecast errors compared to the base model. In the previous analysis and in figure 2 the negative differences were during the period 21:00 till 07:00, whilst the negativity in figure 4 starts at 19:00 till 05:00, showing that the 4 intra-day period model performs better on average compared to the peak/off-peak hour model.

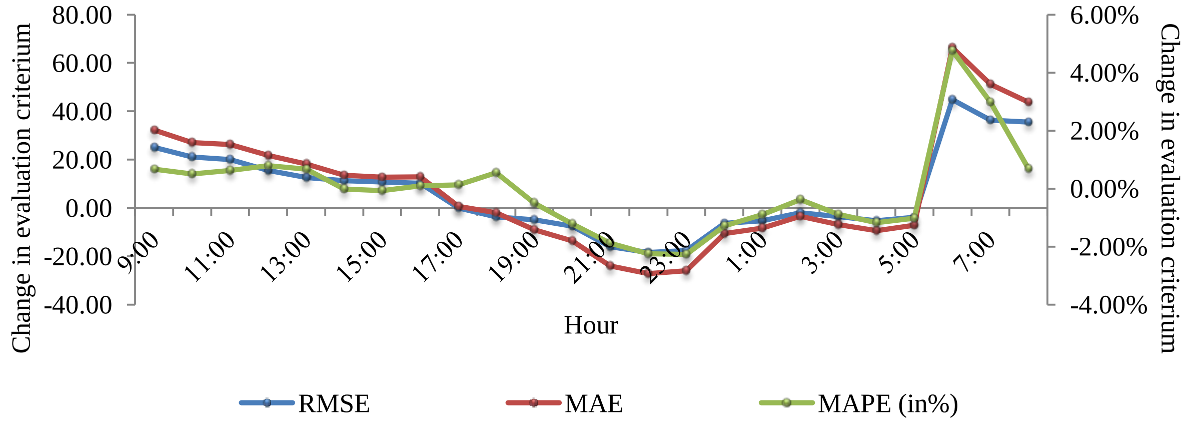

Figure 4: This graph displays the difference in the forecast error evaluation criteria between the 4 intra-day period model and the base model over the period December 30, 2009- March 9, 2010.

Consecutively it can be stated that taking into consideration the different cycles and patterns of the intra-day prices, and slicing the 24 hours into evenly distributed periods, the forecasting performance improves. However, as in the previous analysis of figure 2 the forecast were better for the off-peak hours and worse for the peak hours. Although, the alternative model performs in the current case better, the model cannot capture the variability of the peak hour electricity prices.

The Diebold-Mariano test statistic takes on a value of -1.85, with a corresponding p-value of 0.06 indicating that the forecasts based on the 4 combined intra-day periods forecasts do not generate significantly better forecasts than the base forecast when considering all the hours of the day, at a 5 percent significance level.

The optimal number of principal components for each block can be found in Appendix A-table A4.

4.4 Forecasts combined weighted base and weighted peak/off-peak model

An illustration of the differences in forecast evaluation criteria between the base model and the combined weighted base and weighted peak/off-peak model are presented in figure 5. It strikes to the attention that as opposed to the previous findings in section 4.2 and 4.3, the forecasts obtained from the combined weighted base and weighted peak/off-peak model outperform the plain base model. The differences of the RMSE and MAE are negative between 7:00 and 15:00, which corresponds to peak hours. The positive differences now ranges between 20:00 and 6:00, corresponding with the off peak hours. As opposed to the previous two forecasts from section 4.2 and 4.3 it can be stated that more accurate forecasts for the peak hours are achieved than those arising from the base model for the peak hours. This is, however at the expense of the accuracy of the forecasts of the off-peak hours, as the positive differences are of a higher magnitude.

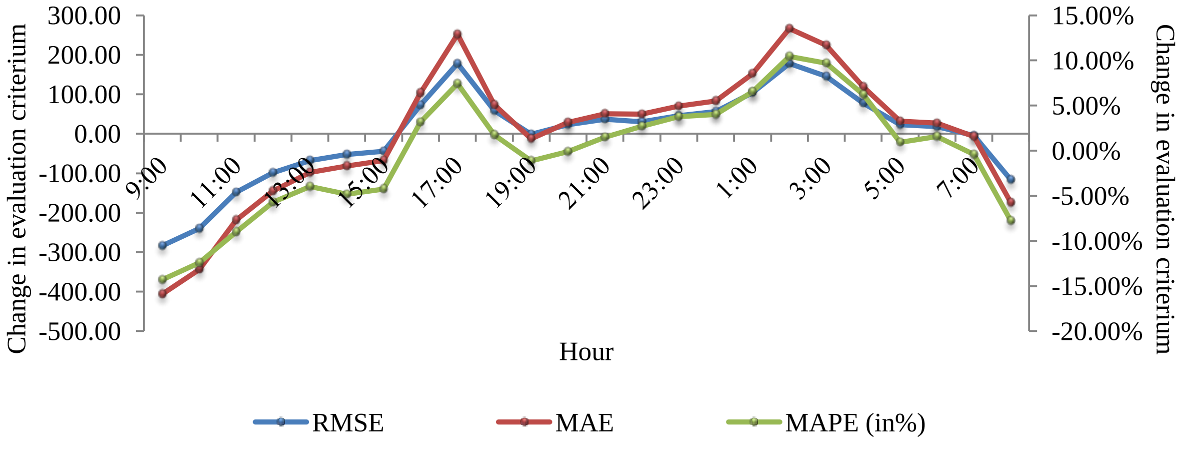

Figure 5: This graph displays the difference in the forecast error evaluation criteria between the combined weighted base and weighted peak/off-peak model and the base model over the period December 30, 2009- March 9, 2010.

The Diebold Mariano test statistic takes on a value of -0.68, with a corresponding p-value of 0.50 indicating that the forecasts based on this weighing alternative model also do not generate significantly better forecasts than the base model at a 5 percent significance level.

4.5 Forecasts weighted principal components

The differences in RMSE, MAE AND MAPE between the results of the weighted principal components forecast and the base forecast are displayed in figure 6. It is immediately clear that the differences between the two forecast methods are not only large but also positive during most hours of the day, indicating that the performance of the weighted principal components model is bad. Very large peaks are observable between the hours 09:00 and 21:00, i.e. the errors from the weighted forecasts are larger than the base model forecasts. The largest difference occurs at 19:00 and is approximately 140.00. Again the differences are negative during the off-peak hours from 21:00 till 07:00.

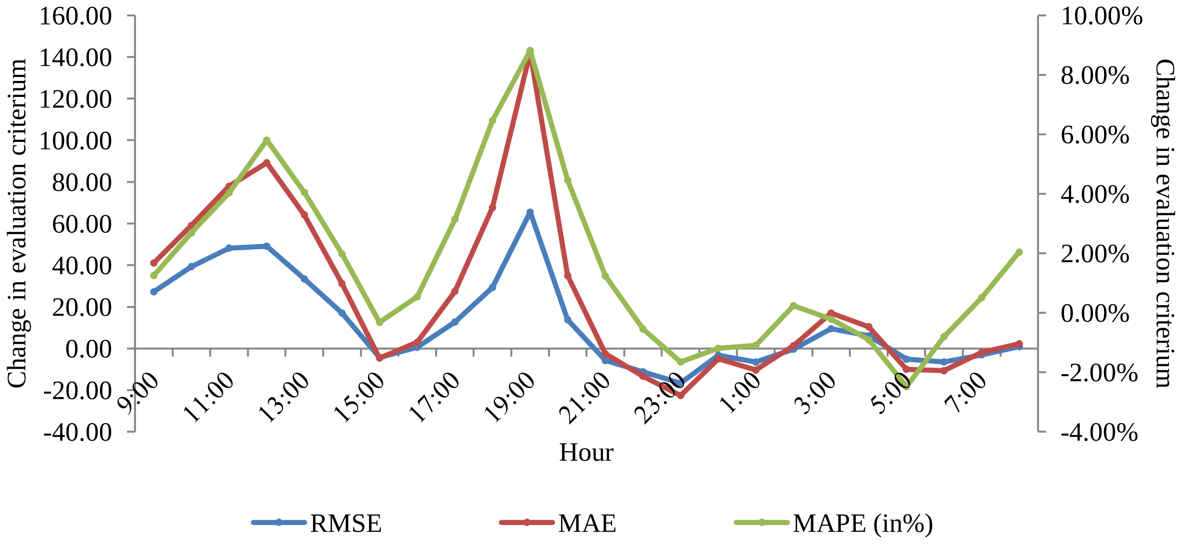

Figure 6: This graph displays the difference in the forecast error evaluation criteria between the combined weighted principal components model and the base model over the period December 30, 2009- March 9, 2010.

The optimal number of principal components for the combined weighted forecasts – when all combinations of forecasts have been considered for a maximum amount of 12 components - has been set to 6 components for the first forecast and 10 components for the second forecast, for the RMSE and MAE. For the MAPE optimal results required slightly different components, namely 5 for the first forecast and 10 for the second forecast. Detailed results, regarding the optimal combinations of weighted principal components can be found in Appendix E-tables E1-E3.

The Diebold-Mariano test statistic takes on a value of -1.34 for a combination of 6 and 10 components and -1.46, for a combination of 5 with 10 components. The corresponding p-values are respectively 0.18 and 0.14 indicating that the both forecasts based on the combined weighted principal components do not generate significantly better forecasts than the base model, at a significance level of 5 percent.

5. Conclusion

Throughout the research comparisons between different models were made. For every comparison the base model i.e.; PCA on 24 hours was used as standard evaluation model. Alternative models which were implemented are the peak/off-peak model, the 4 intra-day periods model, the combined weighted base and weighted peak/off-peak model and the weighted PCA model. The goal was to observe whether these alternative models could improve the accuracy of 1 day-ahead hourly electricity price forecasts. The forecasts have been compared using the DM-statistic.

When comparing the base model with the peak/off-peak model the conclusion can be drawn that on average the latter performs best when applied to the off-peak hour period, in which electricity prices behave less volatile compared to peak hours. The challenge is in making as accurate as possible forecast concerning the peak hours; it is of course during this period when electricity demand is high and with that also the prices. Forecasts obtained from the peak/off-peak hour model did not provide us with significantly better forecasts than the base model at a 5 percent significance level.

By making use of the 4 intra-day period, forecast performance was indeed improved but once again for off peak hours only, forecasts were even worse for the peak hours. The model lacked in capturing the variability over the hourly electricity prices during the peak hours. At a 5 percent significance level, the 4 intra-day period model also failed into providing us with significantly better forecast than the base model.

Forecast accuracy was improved for the peak hours, when considering the difference between the base forecast and the combined weighted base and weighted peak/off-peak forecasts. This however was at the expense of forecast accuracy of the off- peak hours. Overall the model also did not succeed in achieving significantly better results than those from the base model at a 5 percent significance level.

The forecasts obtained from the weighted model compared to those from the base model yielded by far the largest differences in evaluation criteria, concluding that the weighted forecasts model was completely outperformed compared to alternative models. Needless to say that this model also did not provide significantly better forecast at a 5 percent significance level.

It general it can be concluded that it is problematic and difficult to achieve accurate forecast on hourly electricity prices, after data reduction by means of PCA. This is especially the case when it concerns peak hours, which are characterized by high volatility. This seemed to be a common issue among the models examined throughout the research, excluding the forecast obtained from the combined weighted base and weighted peak/off-peak model, but then again, at the expense of forecast accuracy of the off- peak hours. Besides it should not be overlooked that all alternative models have been compared to the so called base model, which in its turn, also does not generate the most excellent forecasting results. Summarizing the findings, a general conclusion can be drawn, stating that after data reduction reasonable forecast results on hourly electricity prices can be achieved, under the condition that volatility is not excessively present, as a principal component regression lacks in capturing dynamics.

6. References

Bai, J., & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 191-221.

Bates, J., & Granger, C. (1969). The Combination of Forecasts. *Operational Research Quarterly*, 451-468.

Chow, T., & Leung, C. (1996, November). Neural network based short-term load forecasting using weather compensation. *IEEE Transactions on Power Systems*, pp. 1736-1742.

Damme, E. v. (2005). *Liberalizing the Dutch Electricity Market: 1998-2004.* Tilburg .

Hannes, K., & Jurgen, P. (2009). A Corrected Criterion for Selecting the Optimum Number of Principal Components. *Austrian Journal of Statistics Volume 38*, 135-150.

Hildmann, M., & Herzog, F. (2010). *Two Empirical Methods for Constructing Price Forward Curves and Forecasting Spot Prices in the Swiss Power Market.* Zurich .

Johnsen, T. A., Verma, S. K., & Wolfram, C. (1999, June 29). POWER. *Zonal Pricing And Demand-Side Bidding In The Norwegian Electricity Market*.

Knittel, C. R., & Roberts, M. R. (2005, May 17). An empirical examination of restructured electricity prices. *Energy Economics, Elsevier*, pp. 791-817.

Mills, T. (1993). *The Econometric Modelling of Financial Time Series* . Cambridge University Press 1999.

Misiorek, A., Trueck, S., & Weron, R. (2007). Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models. *tudies in Nonlinear Dynamics & Econometrics*.

Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 1167-1179.

Stock, J. H., & Watson, M. W. (2004). *Forecasting with many predictors.*

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

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

7. Appendix

Appendix A

Component Number	Eigen Value	Component Number	Eigen Value
1	48600.53	13	29.20
2	6462.94	14	19.88
3	3172.84	15	13.33
4	1070.46	16	12.69
5	761.64	17	10.40
6	440.61	18	10.15
7	387.49	19	6.51
8	328.74	20	6.31
9	195.06	21	4.41
10	73.03	22	3.25
11	41.77	23	2.55
12	38.74	24	1.93

Table A1: The table displays the number of principal components and the respective eigenvalues for PCA on 24 hours, over the estimation period August, 12 2009 – December, 29 2009.

Component Number	Eigen Value	Component Number	Eigen Value
1	37254.59	7	75.74
2	2879.87	8	19.58
3	828.17	9	11.13
4	428.35	10	5.61
5	245.12	11	2.87
6	182.66	12	2.56

Table A2: The table displays the number of principal components and the respective eigenvalues based on PCA on the peak hours, over the estimation period August, 12 2009 – December, 29 2009.

Component Number	Eigen Value	Component Number	Eigen Value
1	16920.80	7	36.04
2	1614.04	8	21.26
3	628.21	9	12.44
4	349.37	10	11.71
5	92.98	11	6.70
6	59.57	12	5.06

Table A3: The table displays the number of principal components and the respective eigenvalues based on PCA on the off-peak hours, over the estimation period August, 12 2009 – December, 29 2009.

Block 1		Block 2		Block 3		Block 4	
Component number	Eigen Value	Component number	Eigen Value	Component number	Eigen Value	Component number	Eigen Value
1	9804.24	1	5517.28	1	20176.42	1	21117.33
2	410.74	2	303.30	2	1298.22	2	1814.17
3	86.62	3	41.04	3	116.15	3	384.11
4	19.73	n/a	n/a	4	78.49	4	309.78
5	11.45	n/a	n/a	5	52.30	5	84.84
n/a	n/a	n/a	n/a	6	7.22	6	40.43
n/a	n/a	n/a	n/a	7	3.43	7	8.90
n/a	n/a	n/a	n/a	8	3.09	8	5.16

Table A4: The table displays the number of principal components and the respective eigenvalues based on PCA on the 4 intra-day periods, over the estimation period August, 12 2009 – December, 29 2009.

Appendix B

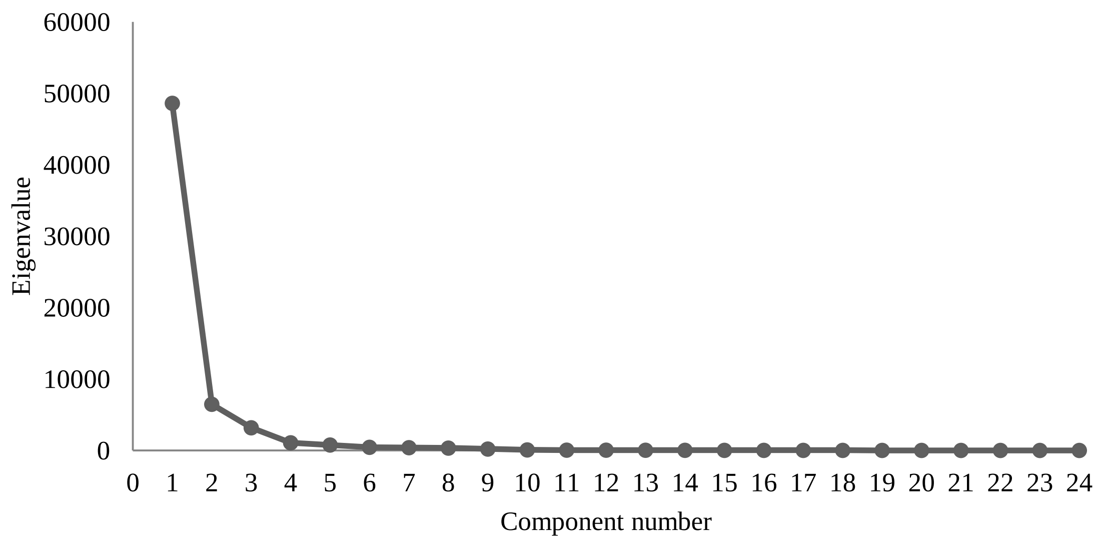

Figure B1: The scree-plot displays the eigenvalues set out against their component number, considering the 24 hours over the estimation period August, 12 2009 – December, 29 2009.

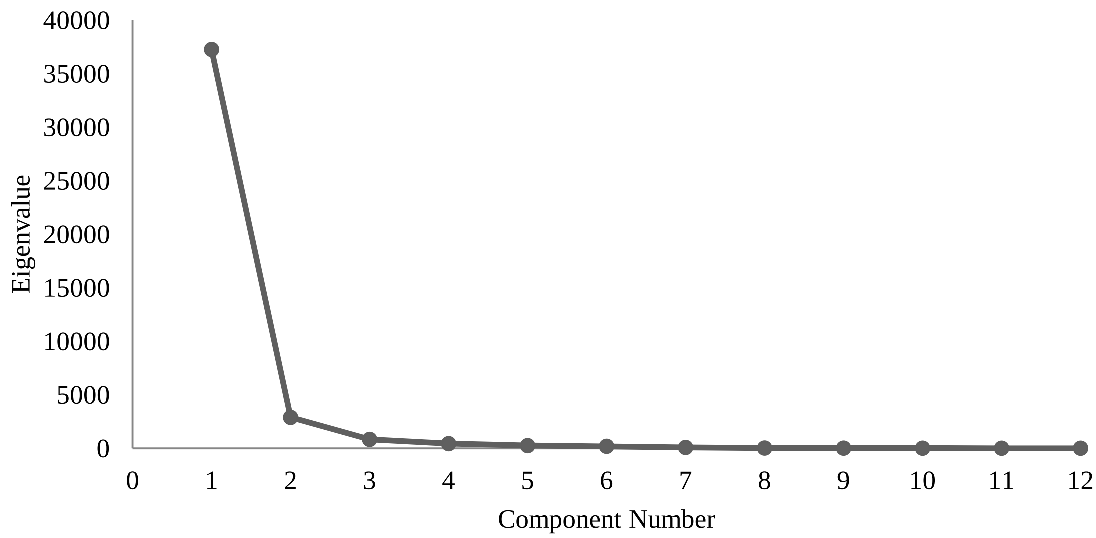

Figure B2: The scree-plot displays the eigenvalues set out against their component number, considering the 12 peak-hours over the estimation period August, 12 2009 – December, 29 2009.

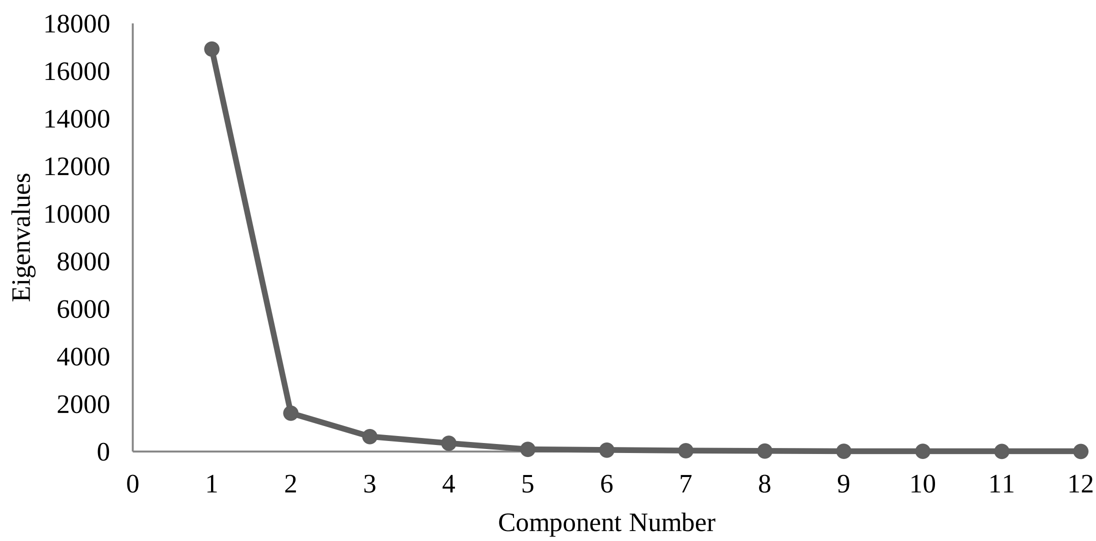

Figure B3: The scree-plot displays the eigenvalues set out against their component number, considering the 12 off-peak hours over the estimation period August, 12 2009 – December, 29 2009.

Appendix C

Hour	Mean	Max	Min	Std. dev	Skewness	Kurtosis
1	263.18	330.56	162.13	36.93	-0.54	2.51
2	253.92	318.29	116.98	40.10	-0.63	2.79
3	247.32	538.94	85.50	51.40	0.75	9.34
4	240.56	313.84	77.39	49.35	-0.70	2.74
5	244.57	318.47	78.91	47.79	-0.68	2.88
6	258.53	325.56	87.94	43.40	-0.86	3.75
7	277.80	350.76	90.48	42.29	-1.43	6.08
8	297.71	434.00	101.62	46.79	-0.85	5.74
9	313.97	805.04	150.74	66.51	3.39	25.87
10	315.22	737.32	173.78	62.51	3.84	27.36
11	314.80	651.46	192.68	51.17	3.06	21.60
12	314.26	628.63	197.49	49.26	2.87	20.11
13	308.63	527.18	194.46	39.13	0.87	9.74
14	304.98	600.06	184.75	42.98	1.83	17.93
15	301.86	651.46	177.07	47.77	2.41	22.35
16	301.71	651.29	173.78	51.95	2.11	16.88
17	311.24	882.49	176.56	77.16	3.69	25.26
18	326.09	1042.32	193.19	97.70	3.99	25.69
19	317.89	823.34	219.60	60.69	4.19	35.72
20	310.50	457.17	228.13	31.94	0.49	5.26
21	302.68	394.90	227.11	26.21	-0.23	3.98
22	295.02	353.87	203.72	27.53	-0.75	3.72
23	286.79	340.27	196.86	30.51	-0.74	3.19
24	269.92	329.15	179.08	34.47	-0.59	2.56

Table C1: The table displays the mean, maximum, minimum, standard deviation, skewness and kurtosis of the hourly electricity prices over the estimation period August, 12 2009 – December, 29 2009.

Appendix D

Hour	RMSE	MAE	MAPE	Hour	RMSE	MAE	MAPE
9:00	306.65	138.61	17.72%	21:00	57.95	35.77	7.19%
10:00	259.13	117.14	15.23%	22:00	52.71	31.51	6.66%
11:00	168.63	85.71	12.42%	23:00	47.82	28.85	6.49%
12:00	121.10	63.44	9.86%	0:00	39.14	25.74	5.84%
13:00	93.35	49.74	8.55%	1:00	27.38	18.40	4.08%
14:00	79.55	47.61	8.77%	2:00	21.71	14.45	3.16%
15:00	86.72	49.69	9.51%	3:00	23.49	17.21	3.94%
16:00	89.08	48.81	9.09%	4:00	26.04	19.99	4.74%
17:00	129.16	64.13	10.22%	5:00	28.25	22.17	5.14%
18:00	199.57	101.79	13.43%	6:00	27.81	20.02	4.21%
19:00	169.26	96.45	13.60%	7:00	43.79	27.99	5.71%
20:00	97.39	58.03	10.06%	8:00	150.59	79.17	12.56%

Table D1: the table displays values for the RMSE, MAE and MAPE for the forecasts obtained from the base model for the period December 30, 2009-March, 9 2010.

Hour	RMSE	MAE	MAPE	Hour	RMSE	MAE	MAPE
9:00	307.01	137.81	17.66%	21:00	50.82	35.11	6.83%
10:00	258.01	116.26	15.22%	22:00	45.90	32.73	6.50%
11:00	167.49	85.73	12.61%	23:00	39.36	28.13	5.92%
12:00	120.47	64.19	10.20%	0:00	33.75	23.21	5.01%
13:00	94.42	53.23	9.20%	1:00	23.13	15.82	3.34%
14:00	83.17	54.82	10.10%	2:00	19.13	12.22	2.61%
15:00	93.26	57.77	11.21%	3:00	20.27	13.22	2.99%
16:00	94.99	58.25	11.03%	4:00	22.40	15.83	3.72%
17:00	131.40	70.10	11.34%	5:00	25.20	19.22	4.29%
18:00	199.35	104.72	13.93%	6:00	26.67	18.92	3.90%
19:00	169.13	98.06	14.05%	7:00	42.19	28.37	5.69%
20:00	103.63	67.70	12.39%	8:00	161.41	83.93	12.67%

Table D2: the table displays values for the RMSE, MAE and MAPE for the forecasts obtained from the peak/off-peak model for the period December 30, 2009-March, 9 2010.

Hour	RMSE	MAE	MAPE	Hour	RMSE	MAE	MAPE
9:00	331.71	145.80	18.39%	21:00	41.91	27.76	5.30%
10:00	280.26	123.09	15.73%	22:00	34.29	22.70	4.41%
11:00	188.63	91.90	13.05%	23:00	30.12	20.58	4.22%
12:00	136.62	69.64	10.65%	0:00	32.72	21.59	4.55%
13:00	105.99	55.23	9.22%	1:00	22.08	15.31	3.19%
14:00	90.84	49.80	8.76%	2:00	19.71	12.98	2.79%
15:00	97.52	51.68	9.44%	3:00	19.88	13.93	3.05%
16:00	99.31	51.45	9.18%	4:00	20.78	15.79	3.56%
17:00	129.10	64.90	10.35%	5:00	24.37	18.83	4.12%
18:00	195.91	103.43	13.97%	6:00	72.63	41.49	8.97%
19:00	164.41	92.29	13.10%	7:00	80.17	42.78	8.68%
20:00	89.78	51.92	8.84%	8:00	186.16	87.38	13.24%

Table D3: the table displays values for the RMSE, MAE and MAPE for the forecasts obtained from the 4 intra-day period model for the period December 30, 2009-March, 9 2010.

Hour	RMSE	MAE	MAPE	Hour	RMSE	MAE	MAPE
9:00	23.13	15.95	3.41%	21:00	94.42	50.14	8.68%
10:00	19.13	13.08	2.82%	22:00	83.17	50.80	9.35%
11:00	20.27	14.95	3.40%	23:00	93.26	53.22	10.25%
12:00	22.40	17.43	4.13%	0:00	94.99	52.64	9.88%
13:00	25.20	20.10	4.59%	1:00	131.40	66.28	10.61%
14:00	26.67	19.11	3.96%	2:00	199.35	102.90	13.62%
15:00	42.19	26.42	5.26%	3:00	169.13	96.20	13.64%
16:00	161.41	79.73	12.25%	4:00	103.63	61.50	10.97%
17:00	307.01	137.95	17.64%	5:00	50.82	31.16	6.08%
18:00	258.01	116.38	15.17%	6:00	45.90	28.87	5.76%
19:00	167.49	85.52	12.48%	7:00	39.36	24.99	5.30%
20:00	120.47	63.53	9.97%	8:00	33.75	21.98	4.78%

Table D4: the table displays values for the RMSE, MAE and MAPE for the forecasts obtained from the combined weighted base and weighted peak/off-peak model for the period December 30, 2009-March, 9 2010.

Hour	RMSE	MAE	MAPE	Hour	RMSE	MAE	MAPE
9:00	333.89	152.41	18.97%	21:00	52.17	39.01	8.42%
10:00	298.47	136.88	17.92%	22:00	41.51	29.38	6.11%
11:00	216.86	115.38	16.45%	23:00	31.28	22.81	4.84%
12:00	170.21	103.55	15.66%	0:00	35.93	23.95	4.65%
13:00	126.73	80.54	12.60%	1:00	20.86	14.54	2.99%
14:00	96.53	61.78	10.75%	2:00	21.38	16.13	3.40%
15:00	82.46	49.36	9.19%	3:00	32.98	24.70	3.73%
16:00	89.65	51.23	9.62%	4:00	32.13	24.33	3.84%
17:00	141.85	78.88	13.36%	5:00	23.08	17.36	2.66%
18:00	228.85	140.19	19.89%	6:00	21.36	15.82	3.40%
19:00	234.73	173.17	22.42%	7:00	40.77	29.17	6.21%
20:00	111.19	79.14	14.52%	8:00	151.46	80.56	14.60%

Table D5: the table displays values for the RMSE, MAE and MAPE for the forecasts obtained from the weighted principal components model for the period December 30, 2009-March, 9 2010.

Appendix E

		Forecast 1											
	PC	1	2	3	4	5	6	7	8	9	10	11	12
Forecast 2	1	n/a											
	2	180.37											
	3	150.16	117.87										
	4	148.99	117.64	118.64									
	5	154.86	122.30	118.30	127.67								
	6	147.43	118.44	117.82	119.93	115.87							
	7	143.33	114.66	116.63	116.72	113.39	122.01						
	8	143.12	115.38	117.11	117.38	114.13	122.85	127.86					
	9	143.05	115.62	117.12	117.38	114.46	123.15	124.41	119.73				
	10	144.27	114.89	115.95	116.44	110.93	109.85	122.55	121.46	123.57			
	11	141.99	115.67	116.40	116.51	113.11	112.10	121.61	120.09	120.16	121.43		
	12	142.02	115.30	116.08	116.34	113.22	111.65	121.46	120.12	120.12	120.66	121.01	n/a

Table E1: the table displays the average RMSE value obtained from the weigted PCA model for the period December 30, 2009-March, 9 2010.

Forecast 1

	PC	1	2	3	4	5	6	7	8	9	10	11	12
	1	n/a											
	2	99.02											
	3	85.00	70.27										
	4	84.54	70.12	70.71									
Forecast 2	5	87.58	71.62	70.22	75.07								
	6	84.28	70.36	69.91	71.71	69.32							
	7	81.18	67.23	68.75	68.56	66.96	74.24						
	8	81.14	67.63	69.09	69.11	67.53	74.43	77.62					
	9	81.10	67.78	69.24	69.16	67.86	75.01	75.96	74.36				
	10	81.69	67.64	68.44	68.85	65.31	65.01	77.31	75.39	77.51			
	11	80.60	68.23	68.69	68.90	66.86	66.51	75.61	73.99	74.51	73.73		
	12	80.65	68.24	68.64	69.06	67.14	66.58	75.68	74.27	74.77	73.09	73.96	n/a

Table E2: the table displays the average MAE value obtained from the weigted PCA model for the period December 30, 2009-March, 9 2010.

Forecast 1

	PC	1	2	3	4	5	6	7	8	9	10	11	12
	1	n/a											
	2	17.64%											
	3	15.41%	11.27%										
	4	15.31%	11.30%	11.31%									
Forecast 2	5	15.94%	11.64%	11.23%	12.30%								
	6	15.19%	11.31%	11.15%	11.46%	10.95%							
	7	14.59%	10.83%	10.92%	10.87%	10.52%	11.90%						
	8	14.55%	10.92%	11.00%	11.00%	10.61%	12.02%	12.63%					
	9	14.55%	10.95%	11.01%	10.99%	10.66%	12.07%	12.20%	11.84%				
	10	14.74%	10.94%	10.93%	10.98%	10.26%	10.33%	12.48%	11.99%	12.41%			
	11	14.44%	11.02%	10.95%	10.98%	10.51%	10.57%	12.16%	11.73%	11.85%	11.66%		
	12	14.48%	11.02%	10.94%	11.02%	10.56%	10.61%	12.17%	11.80%	11.91%	11.52%	11.71%	n/a

Table E3: the table displays the average MAPE value obtained from the weigted PCA model for the period December 30, 2009 March, 9 2010.