Politechnika Wrocławska

Wydział Zarządzania

PCA Forecast Averaging – Predicting Day Ahead Electricity Price Forecasts Across Calibration Windows

By

Hammad Ullah

Supervisor:

Dr. Tomasz Serafin

Contents

1. Introduction

Off lately, we have observed energy markets undergoing a dynamic transformation, which have changes in the generation structure and the creation of new trading opportunities [1]. Ever since the competitive power exchange was established, a growing share of electricity was traded in day-ahead markets, where offers are place pre noon of the day preceding the delivery [1]. Many market participants find the operations in such a complex environment challenging, since it requires taking various operational decisions on how to structure the intraday trade for example. Hence, an accurate prediction of electricity price is of major importance to the utility managers [1].

Typically, the day-ahead electricity price forecasting studies focus on developing model structures that better represent the temporal, inter variable dependencies, input variable selection and finding optimal weights for combined forecasts [2]. But not many studies in the electricity price forecasting try to find the average forecasts obtained models which are estimated from different windows or find the optimal calibration length [2]. Some researchers argue that forecasting performance is sensitive to the calibration window choice, and it might be better to combine forecasts of different lengths, in the presence of structural breaks [2]. Shorter windows often adopt better to changes and longer windows often lead to a more precise estimation of model parameters. Therefore, forecasts obtained from different windows will address distinct features of the underlying process [2].

Principle component analysis (PCA) is a dimensionality reduction method that is often used to reduce the dimensionality of large datasets by transforming a large set of variables into a smaller one that still contains most of the information in the larger set. Reducing the number of variables of a dataset naturally has its disadvantages in terms of accuracy, but advantage is the simplicity due to dimensionality reduction. Due to this, smaller dataset is easier to explore and visualize. Various literatures have reported that the electricity price forecasting using PCA have resulted in more accurate short and midterm forecasts, due to the joint exploration of the forecast panel. The PCA approach could be also used to combine forecasts obtained from different models or/and model specifications [1]. The PCA method have been extensively used and it has been successful for modelling large panels of data.

While some authors try to develop weighting schemes that are optimal, the general conclusion from these studies is that it is hard to outperform and simple arithmetic average across all windows. One can argue that in the context of electricity markets, where the time series of interest (prices, loads) are characterized by weekly and annual seasonal behaviour, that there might be other better alternatives [2]. We try to show in our study, averaging across a few short (e.g., 28, 56 and 84 days) and a few long (e.g., 714, 721 and 728 days) window lengths that it might bring/not, significant accuracy gains.

In our study, for the given dataset i.e., GEFCOM, we present the various forecasting such as single calibration window (Win), simple average window (AW), weighted average window (WAW), expanding weighted average window $(WAW)_{exp}$, PCA forecast combination approach and different calibration window lengths, using an autoregressive model (ARX) to predict the day ahead prices. The aim of our project is to compare the performance of the PCA to the performance of other forecasting models for various calibration window length and chose the model that produces the best prediction for the given dataset.

2. Methodology

The dataset for which the analysis was caried out was the GEFCOM 2014 dataset, for the period 01-01-2011 to 17-12-2013. The dataset contains hourly locational marginal prices (LMP) and day ahead predictions of hourly system load as shown in Fig 1.

We take the 1st 728 days of the dataset as the initial calibration period and 54 days window as PCA, starting from the next day after the end of the initial calibration widow. After the PCA widow, the remaining 300 days of the dataset are considered as test period, as shown in Figure 1. We could observe from Figure 1 that the locational marginal pricing data is leptokurtic in nature, due to the heavy tails caused by the positive spikes.

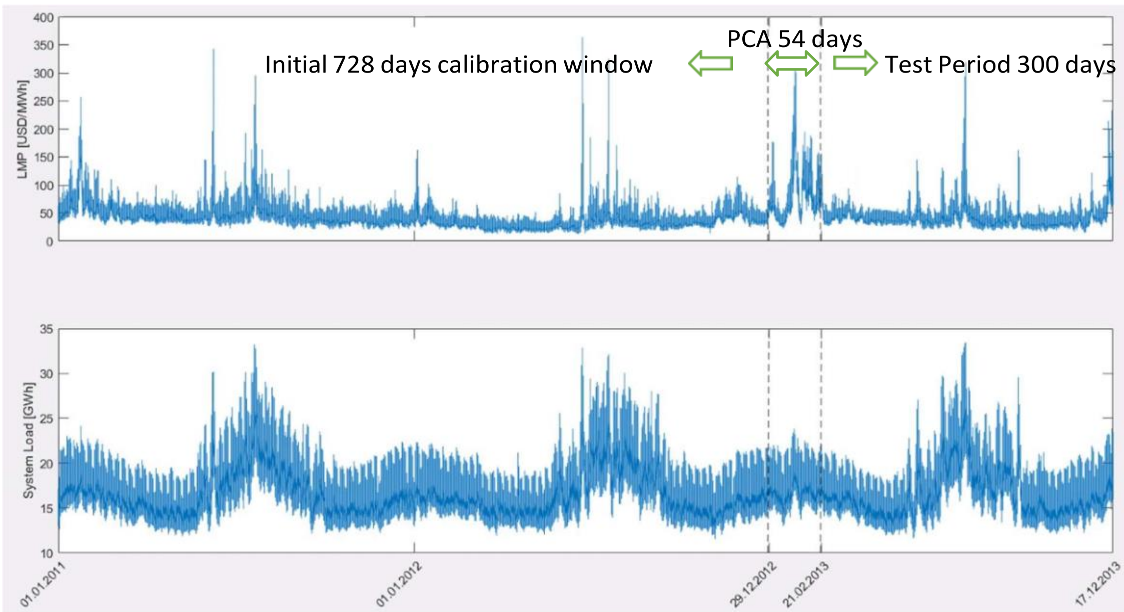

Figure 1: Hourly locational marginal pricing (Top) and System Load (Bottom) of the GEFCOM 2014 dataset

For modelling, we use the 'hour x day' matrix like structure with Eq 1 representing the log price for day d and hour h, for a model for each hour of the day.

Equation 1

$$p_{d,h} \equiv \log(P_{d,h})$$

A well-known structure of the Auto Regressive Model (ARX) was used in our study, as shown in Eq 2.

Equation 2

$$p_{d,h} = \beta_{h,0} + \beta_{h,1}p_{d-1,h} + \beta_{h,2}p_{d-2,h} + \beta_{h,3}p_{d-7,h} + \beta_{h,4}p_{\min} + \beta_{h,5}z_t + \sum_{i\in\{1,6,7\}} \beta_{h,i+5}D_i + \varepsilon_{d,h}$$

Where, $p_{\min} = min_{\mathrm{h}} \{p_{\mathrm{d-1,h}}\}$ creates a link with all yesterday's prices, z_{t} is (the logarithm of) the day-ahead load forecast, Di is the dummy variable for day-of-the-week i and εd,h is the noise.

For the calibration window, we use a rolling window scheme where the data could span from T = 28 to 728 days. The length of the largest calibration window is limited to the 1st 728 days of the data sample. The model is initially calibrated to 728 days I.e., 1.01.2011–28.12.2012 and forecasts for all 24 h of 29.12.2012 are determined. Next, the windows are rolled forward by one day and forecasts for all 24 h of 30.12.2012 are computed.

We use “Win(T)” to determine a calibration window, where Win refers to the simple rolling calibration and T refers to the calibration period length.

We use “AW(T)” to determine a simple average forecast across a set of windows, where we use T = (28,728) refers to 28 days rolling calibration window and 728 days rolling calibration window. We also use T = (28:728) to refer an initial 28 days rolling calibration window which increments until 728 days. Similarly, we use T = (28:28:726) to refer an initial rolling calibration window of 28 days which increment by 28 days and does so until it reaches the max rolling calibration length of 728 days. It is to be noted that the averaging forecast scheme assumes equal weights for all forecasts estimated with the calibration windows.

Similarly, WAW(τ) refers to adding weights to the simple average forecast across a set of windows using the inverse of the mean absolute error (MEA) calculated over the average window length.

$$w_d^{(\tau)} = \frac{\frac{1}{\text{MAE}^{(\tau)}_{d-D_{ave}:d-1}}}{\sum_{v\in\mathcal{T}} \frac{1}{\text{MAE}^{(v)}_{d-D_{ave}:d-1}}}$$

Equation 3

where w^{c}_{d} is the weight corresponding to a window of length τ on day d. $MAE^{(c)}_{d1:d2}$ is calculated as an average of absolute forecast errors obtained from the calibrated model to a training sample of length τ over all hours of days $d \in \{d_1, d_1 + 1, \dots, d_2\}$:

$$\text{MAE}^{(\tau)}_{d_1:d_2} = \frac{1}{24(d_2 - d_1 + 1)} \sum_{d=d_1}^{d_2} \sum_{h=1}^{24} |\varepsilon^{(\tau)}_{d,h}|$$

Equation 4

where $\varepsilon^{(c)}_{d,h} = P_{d,h} - P'^{(c)}_{d,h}$.

WAW(T)^{exp} refers to the expanding weighted average, where at any given day, we take all the days in the past out of sample and that would be our performance window.

To utilize the information included in the panel of forecasts from the above mentioned calibration window types, we made use of principle component analysis (PCA). Here the data and predictions are treated as time series, with the time index, t = 24(d − 1) + h, which represents consecutive hours. As with the WAW method, we use the information from D_{ave}+24 observations. The prediction of the variable P_t based on a calibration window of the length τ, is denoted by $P'_{t,c}$ where the first dimension represents time, and the latter represents the size of the calibration window.

Steps for the PCA algorithm are as follows:

- For each time period t in the averaging window, we estimate the mean (μˆt) and standard deviation (σˆt) of individual forecasts across different τ.
- We standardize/normalize the predictions and predicted variables using equation 5.

$$\tilde{P}_{t,\tau} = \frac{\hat{P}_{t,\tau} - \hat{\mu}_t}{\hat{\sigma}_t}$$

Equation 5

- We then estimate the first k = 1, ..., K principal components, PCt,k , of a panel {P˜ t,τ}. We denote the model by PCA(K), based on the number of principal component K used in our model.

- We then run a regression using observations from the averaging window, without the last 24 observations using equation 6

$$\tilde{P}_t = \alpha + \sum_{k=1}^{K} \beta_k PC_{t,k} + \varepsilon_t$$

Equation 6

- Finally, we calculate the normalized dependant variable on day d and hour h and we then transform it into original units using the following equations below.

$$\hat{\tilde{P}}_{24(d-1)+h} = \hat{\alpha} + \sum_{k=1}^{K} \hat{\beta}_k PC_{24(d-1)+h,k}$$

Equation 7

$$\hat{P}_{24(d-1)+h} = \hat{\tilde{P}}_{24(d-1)+h} * \hat{\sigma}_{24(d-1)+h} + \hat{\mu}_{24(d-1)+h}$$

Equation 8

Some of the calibration window lengths for average forecasts and weighted forecasts across a set of windows and PCAs, for which the MAE were calculated in our study are as follows:

Table 1: Various calibration types and window lengths

Calibration Window	Calibration Window	Calibration Window
Win(28)	WAW(28:14:728)	WAW(28,56,728) exp
Win(364)	AW(28:28:728)	WAW(28,56,728)
AW(364,728)	WAW(28:28:728) exp	AW(28,56,364,728)
WAW(364,728) exp	WAW(28:28:728)	WAW(28,56,364,728) exp
WAW(364,728)	AW(56,728)	WAW(28,56,364,728)
Win(728)	WAW(56,728) exp	AW(28,56,721,728)
AW(28:728)	WAW(56,728)	WAW(28,56,721,728) exp
WAW(28:728) exp	AW(28,728)	WAW(28,56,721,728)
WAW(28:728)	WAW(28,728) exp	PCA(1)
AW(28:7:728)	WAW(28,728)	PCA(2)
WAW(28:7:728) exp	AW(28:28:84,714:7:728)	PCA(3)
WAW(28:7:728)	WAW(28:28:84,714:7:728) exp	PCA(4)
AW(28:14:728)	WAW(28:28:84,714:7:728)	
WAW(28:14:728) exp	AW(28,56,728)	

The forecasts were evaluated in terms of Mean Absolute Error (MAE) for the test period of 300 days I.e., 21-12-2012 to 17-12-2013, using the equation below.

$$\mathrm{MAE} = \frac{1}{24D} \sum_{d=1}^{D} \sum_{h=1}^{24} \left|\widehat{\varepsilon}_{d,h}\right|.$$

Equation 9

Finally, we plot the MAE as the error of the length of the calibration windows for simple average forecasts, weighted average forecasts, expanding weighted average forecast across a set of windows and the PCA for all 300 days (test period). The calibration window type which produces the highest positive change would be chosen as the best result compared to the reference calibration widow for the dataset provided. To obtain a clear winner among all the models in our study, we calculated the “Percentage Change” of all the models by comparing it to the reference window.

3. Results

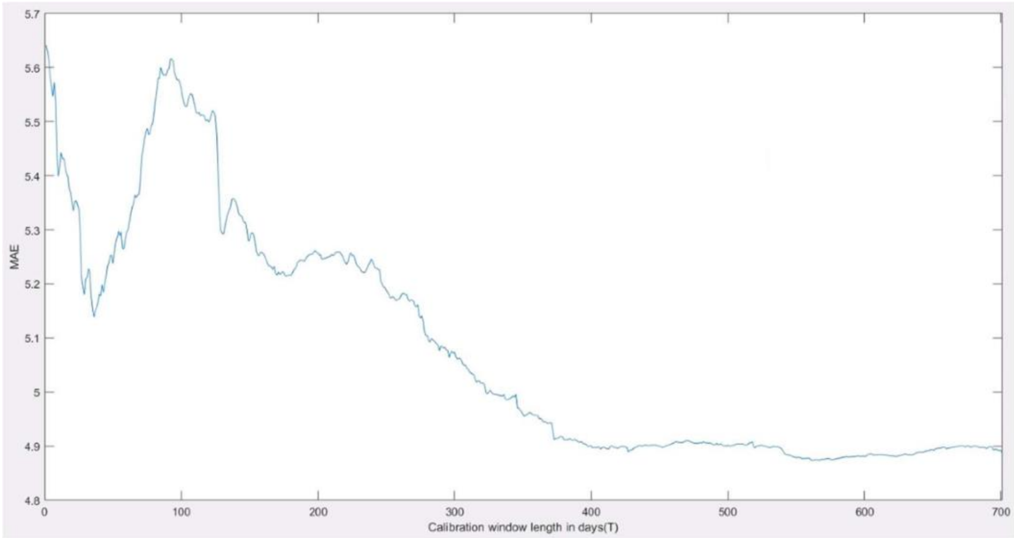

Figure 2: Rolling calibration window length in days from smallest to largest

Figure 2 shows the mean Absolute Errors (MAE) as a function of the window length T = 28,..., 728 (blue circles) and obtained by averaging forecasts across calibration windows (lines with symbols representing window lengths) for the ARX model.

From table 2 we can observe that most of the forecasting model with various calibration window length outperforms the PCA forecasting. Weighted average window(28,56,721,728) which gave us a positive 6.39% change, performed the best compared to any other models for the dataset used in our study.

Table 2: MEA results and %changes for various forecasting schemes and window length

Windows	MAE	%change
Win(28)	5.6412	-13.40%
Win(364)	4.9879	-2.05%
AW(364,728)	4.9163	-0.63%
WAW(364,728) exp	4.8452	0.83%
WAW(364,728)	4.8448	0.84%
Win(728)	4.8854	0.00%
AW(28:728)	4.8517	0.69%
WAW(28:728) exp	4.792	1.95%
WAW(28:728)	4.7827	2.15%
AW(28:7:728)	4.8456	0.82%
WAW(28:7:728) exp	4.7869	2.06%
WAW(28:7:728)	4.7773	2.26%
AW(28:14:728)	4.836	1.02%
WAW(28:14:728) exp	4.7783	2.24%
WAW(28:14:728)	4.7701	2.42%
AW(28:28:728)	4.8228	1.30%
WAW(28:28:728) exp	4.7677	2.47%
WAW(28:28:728)	4.7601	2.63%

Reference window

AW(56,728)	4.7489	2.87%
WAW(56,728) exp	4.7153	3.61%
WAW(56,728)	4.7071	3.79%
AW(28,728)	4.6892	4.18%
WAW(28,728) exp	4.6586	4.87%
WAW(28,728)	4.6711	4.59%
AW(28:28:84,714:7:728)	4.8288	1.17%
WAW(28:28:84,714:7:728) exp	4.613	5.91%
WAW(28:28:84,714:7:728)	4.5999	6.21%
AW(28,56,728)	4.6739	4.53%
WAW(28,56,728) exp	4.6548	4.95%
WAW(28,56,728)	4.6333	5.44%
AW(28,56,364,728)	4.6142	5.88%
WAW(28,56,364,728) exp	4.5956	6.31%
WAW(28,56,364,728)	4.5957	6.30%
AW(28,56,721,728)	4.6115	5.94%
WAW(28,56,721,728) exp	4.5931	6.36%
WAW(28,56,721,728)	**4.592**	**6.39%**
PCA(1)	4.8707	0.30%
PCA(2)	4.8529	0.67%
PCA(3)	4.773	2.35%
PCA(4)	4.7755	2.30%

Best forecasting scheme and window

A “treemaps” visualization allowed us to better visualize the best performing forecasting model and calibration window length in a hierarchical order, for the given dataset.

Figure 3: Best forecasting scheme and window length by rank

4. Conclusion

Our study concludes that, for the GEFCOM dataset which showed positive spikes of heavy tails, the forecasting scheme and calibration window length which gave the best results was weighted average forecasting I.e., WAW(28,56,721,728). It showed the highest positive percentage change compared to the reference window. Close behind in terms of rank were WAW_{exp}(28,56,721,728) and WAW_{exp}(28,56,364,728). The PCA forecasting did not produce good results compared to the weighted average forecasting for our dataset. One of the reasons for the low performance of PCA could also be due to absence of variance stabilizing transformation (VST) for our dataset in our study. While this was just our opinion, the real reason for the poor performance of the PCA could be due to other unknown reasons. Moreover, increasing the calibration length of the PCA model from 54 days to 100 days to check if it fetched better results for the GEFCOM dataset actually resulted in much poorer results. While we acknowledged that the PCA worked best for the ID3 dataset, it simply wasn't the case for GEFCOM dataset.

References

[1] K. Maciejowska, B. Uniejewski and T. Serafin, "PCA Forecasting Averaging - Prediciting day ahead and intraday electricity prices," *Energies,* 2020.

[2] K. Hubicka, G. Marcjasz and R. Weron, "A note on averaging day ahead electricity price forecasts across calibration windows," *IEEE Transcations on sutainable energy vol 10,* 2019.