Averaging day-ahead electricity price forecasts for autoregressive models across calibration windows of various lengths*

Grzegorz Marcjasz[1,2], Tomasz Serafin[1,2], Rafał Weron[1]

[1]Department of Operations Research, Faculty of Computer Science & Management
and [2]Faculty of Pure and Applied Mathematics
Wrocław University of Science and Technology, Poland

* Supported by the National Science Center (PL) through grants no. 2015/17/B/HS4/00334 and 2016/23/G/HS4/01005

Marcjasz, Serafin & Weron (2018)

energies

MDPI

Article

Selection of Calibration Windows for Day-Ahead Electricity Price Forecasting

Grzegorz Marcjasz [1,2], Tomasz Serafin [1,2] and Rafał Weron [1,*]

1 Department of Operations Research, Faculty of Computer Science and Management, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; gelusz@hotmail.co.uk (G.M.), tomaszserafin.96@gmail.com (T.S.)

2 Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

* Correspondence: rafal.weron@pwr.edu.pl; Tel.: +48-71-320-4525

Received: 17 August 2018; Accepted: 5 September 2018; Published: 7 September 2018

check for updates

Abstract: We conduct an extensive empirical study on the selection of calibration windows for day-ahead electricity price forecasting, which involves six year-long datasets from three major power markets and four autoregressive expert models fitted either to raw or transformed prices. Since the variability of prediction errors across windows of different lengths and across datasets can be substantial, selecting ex-ante one window is risky. Instead, we argue that averaging forecasts

Part 1

Notation and previous results

Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 137 (2007) 134–161
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Selection of estimation window in the presence of breaks

M. Hashem Pesaran[a,*], Allan Timmermann[b]

[a]University of Cambridge, UK
[b]University of Califor

Available onlin

Abstract

In situations where a regression model is subjec
optimal to use pre-break data to estimate the p
sample forecasts. The issue of how best to explo
forecast error variance is explored and illustrated
assumption of strictly exogenous regressors. In pr

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2018.2869557, IEEE Transactions on Sustainable Energy

A Note on Averaging Day-Ahead Electricity Price Forecasts Across Calibration Windows

Katarzyna Hubicka, Grzegorz Marcjasz and Rafał Weron

***Abstract*—We propose a novel concept in energy forecasting and show that averaging day-ahead electricity price forecasts of a predictive model across 28- to 728-day calibration windows yields better results than selecting only one 'optimal' window length. Even more significant accuracy gains can be achieved by averaging over a few, carefully selected windows.**

***Index Terms*—Electricity price forecasting, Combining forecasts, Calibration window, Autoregression, NARX neural network, Committee machine, Diebold-Mariano test**

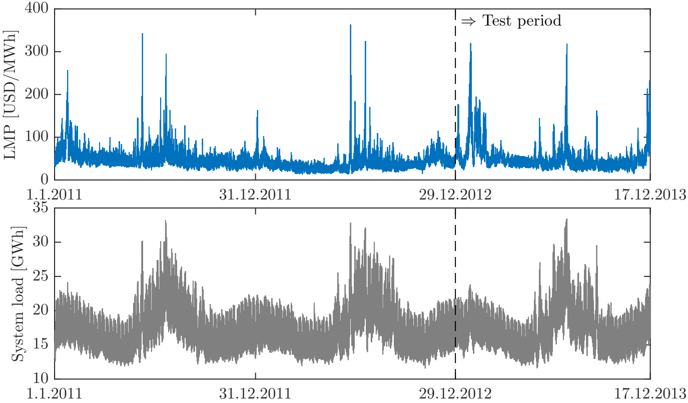
from 28 to 728 days outperforms selecting (even *ex-post*) only one 'optimal' window. Furthermore, we argue that in the context of electricity markets, where the time series of interest (prices, loads) are characterized by weekly and annual seasonal behavior, there may yet be a better alternative. Indeed, as we show below, 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 brings further, significant accuracy gains.

ARX1 (Misiorek et al., 2006, SNDE)

$$X_{d,h} = \beta_{h,0} + \underbrace{\beta_{h,1}X_{d-1,h} + \beta_{h,2}X_{d-2,h} + \beta_{h,3}X_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\beta_{h,4}X_{d-1,min}}_{\text{non-linear effects}}$$

$$+ \underbrace{\beta_{h,5}C_{d,h}}_{\text{load forecast}} + \underbrace{\beta_{h,6}D_{Sat} + \beta_{h,7}D_{Sun} + \beta_{h,8}D_{Mon}}_{\text{weekday dummies}} + \varepsilon_{d,h}$$

GEFCom2014 (01.01.2011-17.12.2013)



Rolling window scheme

- Which window length is optimal? Ten days? Four weeks? A year?
    - No consistency in the EPF literature

Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast

Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting

Grzegorz Dudek
Department of Electrical Engineering, Czesto

ARTICLE INFO

Keywords:
Neural networks
Multilayer perceptron
Probabilistic electricity price forecasting

A Statistical Approach for Interval Forecasting of the Electricity Price

Jun Hua Zhao, *Student Member, IEEE*, Zhao Yang Dong, *Senior Member, IEEE*, Zhao Xu, *Member, IEEE*, and Kit Po Wong, *Fellow, IEEE*

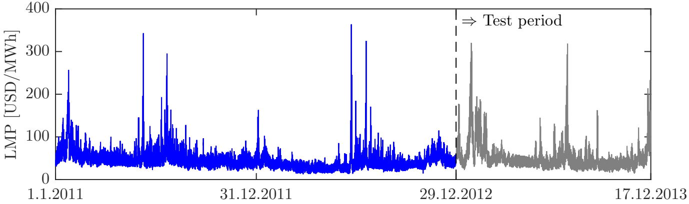
There are two major challenges for accurately electricity price interval forecasting of the electricity price: 1) to estimate the prediction interval, the value of the future price should be accurately forecasted. However, this is difficult because the electricity price is a nonlinear time series, which is highly volatile and cannot be properly modeled by traditional linear

Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast

On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks

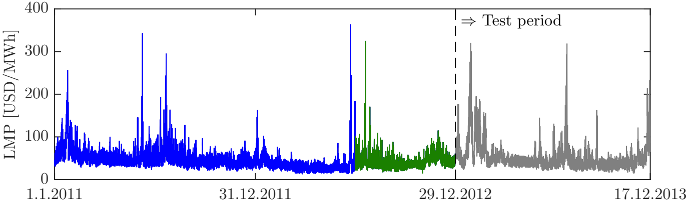
Grzegorz Marcjasz[a,b], Bartosz Uniejewski[a,b], Rafał Weron[a,*]
[a] Department of Operations Research, Wrocław University of Science and Technology, Wrocław, Poland
[b] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland

Rolling window scheme



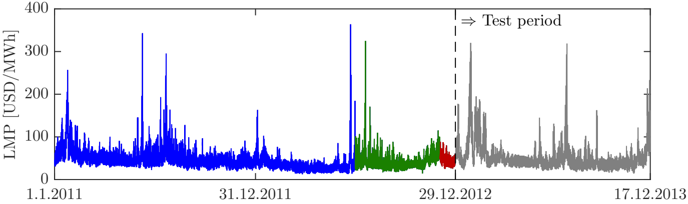
- 728-day calibration window

Rolling window scheme



- 728-day calibration window
- 182-day calibration window

Rolling window scheme

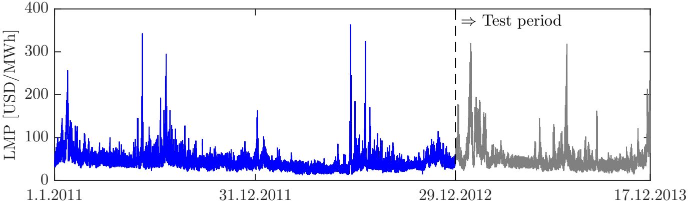


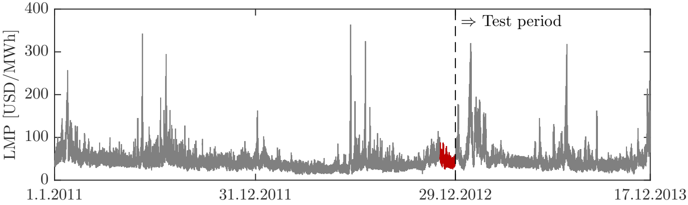
- 728-day calibration window
- 182-day calibration window
- 28-day calibration window

Averaging forecasts: notation

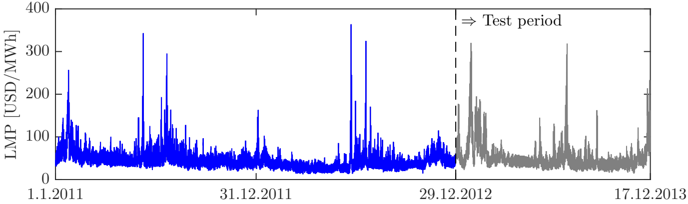
- **Win**(T) - forecast for a T-day window
- **AW**(τ) - simple arithmetic average of $Win(T)$'s
  - $\tau =$ (28,728) refers to 28- and 728-day windows
  - $\tau =$ (28:1092) refers to all 1065 windows ranging from 28 to 1092 days
  - $\tau =$ (28:28:728) is the selection of 26 windows: 28-, 56-, 84-, ..., 728-day
- Mean Absolute Error (MAE)
  - $\text{MAE} = \frac{1}{24D}\sum_{d=1}^{D}\sum_{h=1}^{24}|\widehat{\epsilon}_{d,h}|$,
  - $\widehat{\epsilon}_{d,h} = P_{d,h} - \widehat{P}_{d,h}$ is the forecast error for day d and hour h

Which windows to choose?



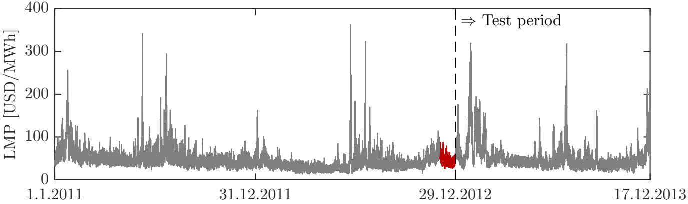


Which windows to choose?



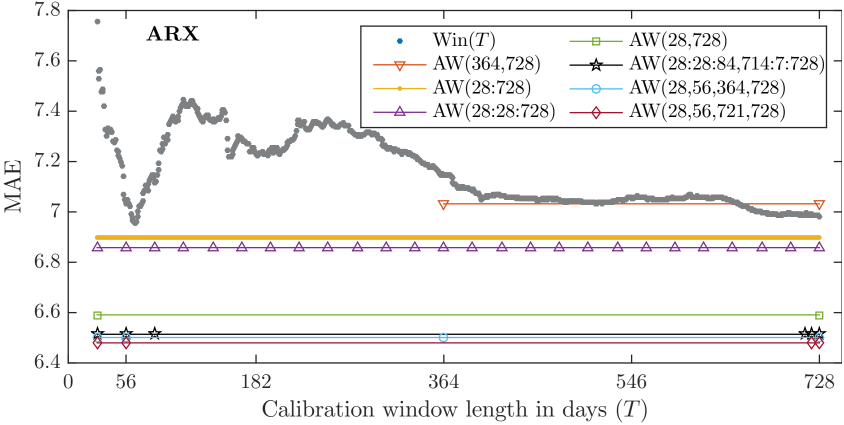
- Longer calibration windows
  - Better reflect trends and allow more precise and stable estimation of model parameters

Which windows to choose?

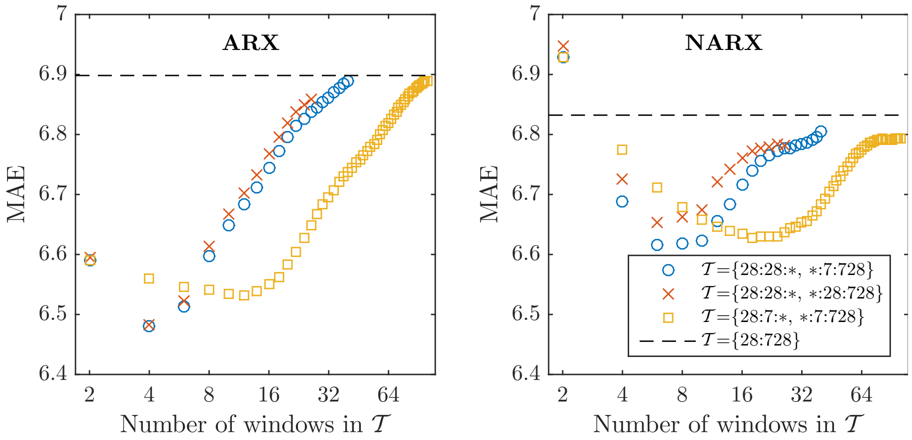


- Shorter calibration windows
  - Tend to quickly adapt to changes in price behavior

Averaging forecasts: results



Averaging forecasts: results



- AW(28:28:84,714:7:728)

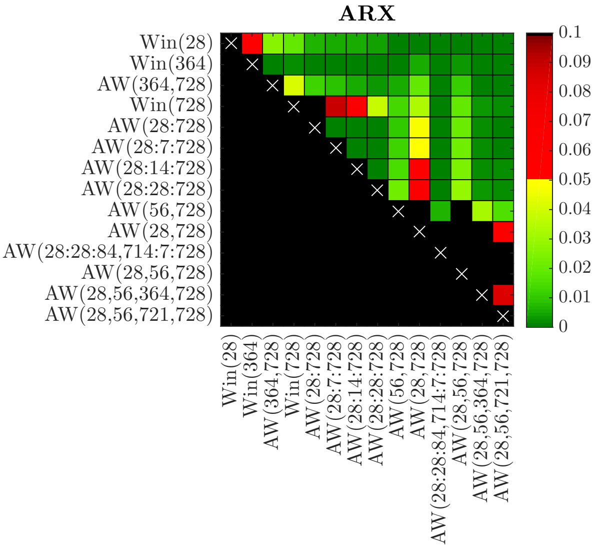
Diebold-Mariano test

- Pair of models: (X,Y)
- Compute 24-dimensional vector of errors for each day

$$\Delta_{X,Y,d} = \|\mathcal{E}_{X,d}\| - \|\mathcal{E}_{Y,d}\|$$

- $\|\mathcal{E}_{X,d}\| = \sum_{h=1}^{24} \widehat{e}_{X,d,h}$
- Hypothesis H_0:$E(\Delta_{X,Y,d}) = 0$
- Hypothesis H_1:$E(\Delta_{X,Y,d}) < 0$
- Hypothesis H_1^{R}:$E(\Delta_{X,Y,d}) > 0$

Diebold-Mariano test: *p*-values



- **Dark green** → *p*-value close to zero → significant difference
- **Black square** → results of X-axis model do not statistically outperform results of Y-axis model

Part 2

Extension of the study

Marcjasz, Serafin & Weron (2018)

- Three recent and much longer datasets (windows: 28-1092 days)
- Models with more explanatory variables
- Data transformation
- Better statistical test
- **New averaging scheme**

energies MDPI

Article

Selection of Calibration Windows for Day-Ahead Electricity Price Forecasting

Grzegorz Marcjasz [1,2], Tomasz Serafin [1,2] and Rafał Weron [1,*]

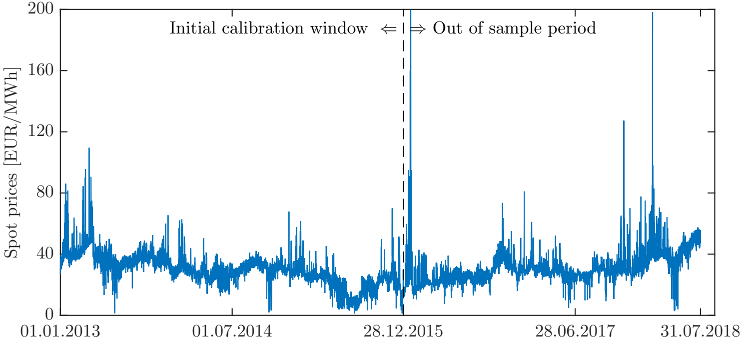
[1] Department of Operations Research, Faculty of Computer Science and Management, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; gelusz@hotmail.co.uk (G.M.), tomaszserafin.96@gmail.com (T.S.)
[2] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, 50-370 Wrocław, Poland
* Correspondence: rafal.weron@pwr.edu.pl; Tel.: +48-71-320-4525

Received: 17 August 2018; Accepted: 5 September 2018; Published: 7 September 2018

check for updates

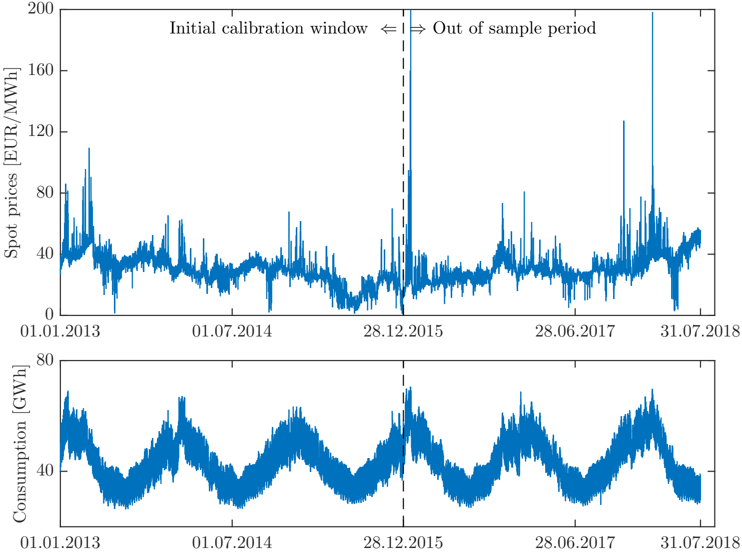
Abstract: We conduct an extensive empirical study on the selection of calibration windows for day-ahead electricity price forecasting, which involves six year-long datasets from three major power markets and four autoregressive expert models fitted either to raw or transformed prices. Since the variability of prediction errors across windows of different lengths and across datasets can be substantial, selecting ex-ante one window is risky. Instead, we argue that averaging forecasts

Dataset: Nord Pool (01.01.2013 - 31.07.2018)

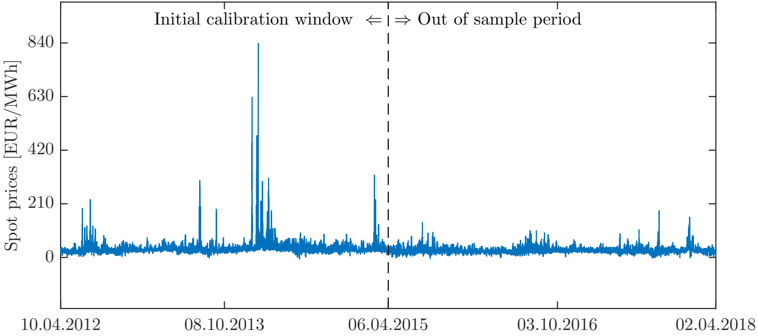


- Hydro-dominated market
- Exhibits strong seasonal variations

Dataset: Nord Pool (01.01.2013 - 31.07.2018)

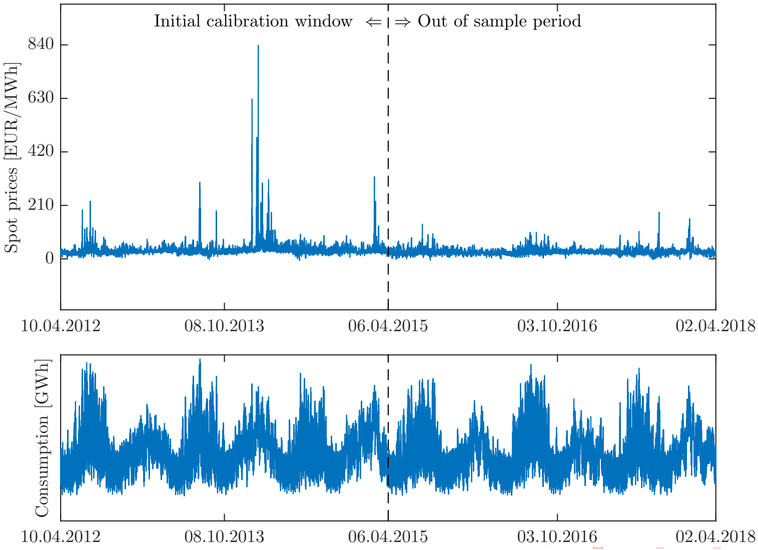


Dataset: PJM (10.04.2012-02.04.2018)

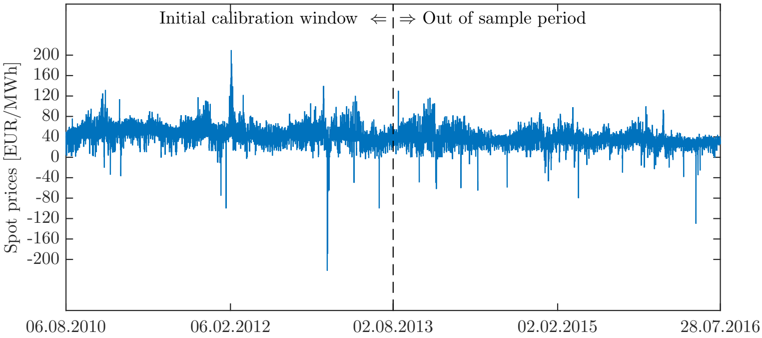


- One of the world's largest wholesale electricity market
- Volatile behavior, particularly in early 2014

Dataset: PJM (10.04.2012-02.04.2018)



Dataset: EPEX (06.08.2010-28.07.2016)



- Rapidly growing share of renewables
- Pronounced negative prices

Variance stabilizing transformation (VST)

IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 33, NO. 2, MARCH 2018 2219

Variance Stabilizing Transformations for Electricity Spot Price Forecasting

Bartosz Uniejewski, Rafał Weron, and Florian Ziel

***Abstract*—Most electricity spot price series exhibit price spikes. These extreme observations may significantly impact the obtained model estimates and hence reduce efficiency of the employed predictive algorithms. For markets with only positive prices, the logarithmic transform is the single most commonly used technique to reduce spike severity and consequently stabilize the variance. However, for datasets with very close to zero (like the Spanish) or negative (like the German) prices the log-transform is not feasible. What reasonable choices do we have then? To address this issue, we evaluate 16 variance stabilizing transformations within a comprehensive forecasting study involving two model classes (regression models, neural networks) and 12 datasets from diverse power markets. We show that the *probability integral transform* combined with the standard Gaussian distribution yields the best approach, significantly better than many of the considered alternatives.**

***Index Terms*—Diebold-mariano test, electricity spot price, forecasting, price spike, probability integral transform, variance stabilizing transformation.**

$Y_{d-1}^{\max}$	Yesterday's maximum VST-transformed price
$\beta_{h,i}$	ith coefficient of the benchmark forecasting model for hour h, defined in (1)
D_i	Weekday dummy for day of the week i
$\varepsilon_{d,h}$	Noise term for day d and hour h
$\hat{F}_Z$	Estimate or distributional forecast of F_Z
$\widehat{\boldsymbol{\varepsilon}}_{X,d}$	Vector of 24 hourly out-of-sample errors for day d of VST X, i.e., $(\widehat{\varepsilon}_{X,d,1}, \ldots, \widehat{\varepsilon}_{X,d,24})'$
$\|\cdot\|_p$	p-norm, i.e., $\|\widehat{\boldsymbol{\varepsilon}}_{X,d}\|_p = (\sum_{h=1}^{24} \|\widehat{\varepsilon}_{X,d,h}\|^p)^{1/p}$
$\Delta_{X,Y,d}$	Loss differential series of the multivariate Diebold-Mariano (DM) test, see (18)
$\Delta_{X,Y,d,M}$	Loss differential series of the multivariate DM test across 11 European markets $(M_1, \ldots, M_{11})$, see (19)

I. Introduction

Variance stabilizing transformation (VST)

- Data 'normalization' prior to applying a VST: $p_{d,h} = \frac{1}{b}(P_{d,h} - a)$
  a is the median of $P_{d,h}$
  b is the sample *median absolute deviation* (MAD)

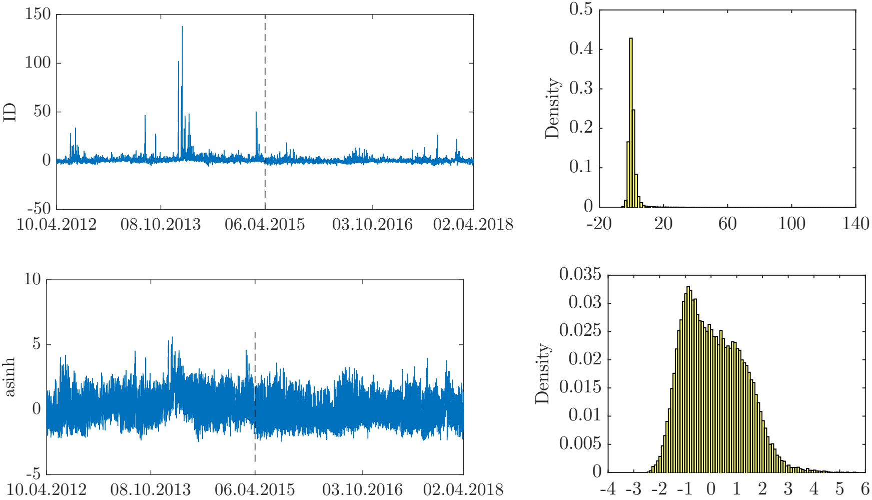
- Area hyperbolic sine (**asinh**)

$$Y_{d,h} = \mathbf{asinh}(p_{d,h}) \equiv \log\left(p_{d,h} + \sqrt{p_{d,h}^2 + 1}\right)$$

- Identity (**ID**)

$$Y_{d,h} = p_{d,h}$$

VSTs: Sample application (PJM dataset)



ARX2 (Weron & Ziel, 2018, Energy Economics)

$$X_{d,h} = \underbrace{\beta_{h,1}X_{d-1,h} + \beta_{h,2}X_{d-2,h} + \beta_{h,3}X_{d-7,h}}_{\text{autoregressive effects}}$$

$$+ \underbrace{\beta_{h,4}X_{d-1,min} + \beta_{h,5}X_{d-1,max}}_{\text{non-linear effects}}$$

$$+ \beta_{h,6}X_{d-1,24} + \underbrace{\beta_{h,7}C_{d,h}}_{\text{load forecast}} + \underbrace{\sum_{i=1}^{7}\beta_{h,7+i}D_i}_{\text{weekday dummies}} + \varepsilon_{d,h}.$$

Weighted Averaged Windows (**WAW**)

- For each window in the combination, calculate MAE over the last t days (e.g. $t = 1$)
- Weight for each window is based on it's past performance over t days:

$$w_T = \frac{\frac{1}{\mathrm{MAE}_{t,T}}}{\sum_{T \in \mathcal{T}} \frac{1}{\mathrm{MAE}_{t,T}}}$$

- Aggregate forecast is a weighted sum of the forecasts for different calibration windows

$$\widehat{P}_{d,h} = \sum_{T \in \mathcal{T}} w_T \widehat{P}_{d,h,T}$$

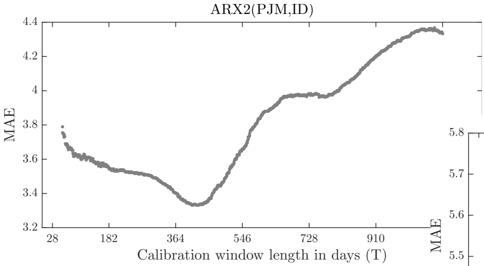
Weighted Averaged Windows (**WAW**): example

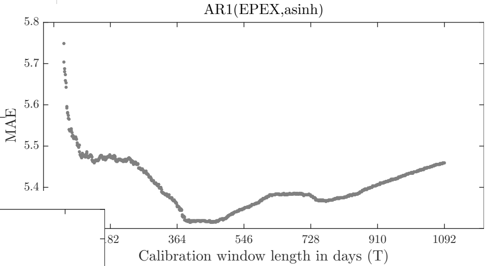
- Window combination: (28,364,728)
- MAE and weights for windows over past 24 hours :

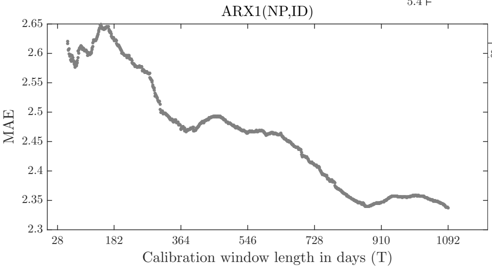
Window	MAE	weight
28	4.2	0.319
364	3.8	0.354
728	4.1	0.327

- Predictions for the next 24 hours are weighted using weights from the table

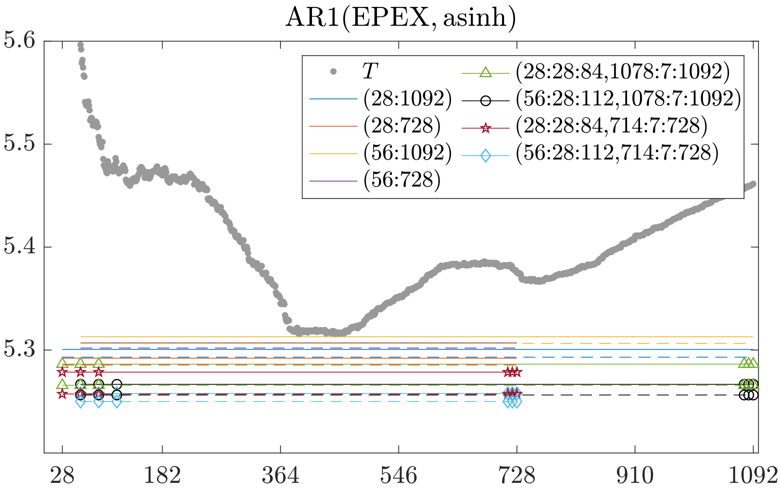
Why should we weigh the forecasts?







Results (— AW , – – WAW)



Results: PJM

Window	ARX1(PJM, ID) Win		ARX1(PJM, asinh) Win		ARX2(PJM, ID) Win		ARX2(PJM, asinh) Win	
28	3.995		68.409		5.141		∞	
56	3.563		3.288		3.691		3.365	
364	3.433		3.196		3.383		3.093	
728	4.078		3.253		3.976		3.121	
1092	4.391		3.294		4.321		3.157	
Window Set	**AW**	**WAW**	**AW**	**WAW**	**AW**	**WAW**	**AW**	**WAW**
(28:1092)	3.653	3.515	3.221	3.178	3.526	3.414	∞	∞
(28:728)	3.465	3.379	3.231	3.178	3.367	3.300	∞	∞
(56:1092)	3.684	3.549	3.170	3.156	3.555	3.443	3.053	3.046
(56:728)	3.499	3.415	3.148	3.134	3.399	3.330	3.042	3.035
(28:28:84,1078:7:1092)	3.563	3.379	13.811	9.853	3.557	3.422	∞	∞
(56:28:112,1078:7:1092)	3.620	3.421	3.090	3.069	3.536	3.380	2.996	2.985
(28:28:84,714:7:728)	3.463	3.335	13.801	10.360	3.458	3.356	∞	∞
(56:28:112,714:7:728)	3.517	3.377	3.080	3.061	3.435	3.321	2.989	2.980

- **WAW**(56:28:112,714:7:728) - 3 out of 4 best forecasts overall

Results: PJM

Window	ARX1(PJM, ID) Win		ARX1(PJM, asinh) Win		ARX2(PJM, ID) Win		ARX2(PJM, asinh) Win	
28	3.995		68.409		5.141		∞	
56	3.563		3.288		3.691		3.365	
364	3.433		3.196		3.383		3.093	
728	4.078		3.253		3.976		3.121	
1092	4.391		3.294		4.321		3.157	
Window Set	**AW**	**WAW**	**AW**	**WAW**	**AW**	**WAW**	**AW**	**WAW**
(28:1092)	3.653	3.515	3.221	3.178	3.526	3.414	∞	∞
(28:728)	3.465	3.379	3.231	3.178	3.367	3.300	∞	∞
(56:1092)	3.684	3.549	3.170	3.156	3.555	3.443	3.053	3.046
(56:728)	3.499	3.415	3.148	3.134	3.399	3.330	3.042	3.035
(28:28:84,1078:7:1092)	3.563	3.379	13.811	9.853	3.557	3.422	∞	∞
(56:28:112,1078:7:1092)	3.620	3.421	3.090	3.069	3.536	3.380	2.996	2.985
(28:28:84,714:7:728)	3.463	3.335	13.801	10.360	3.458	3.356	∞	∞
(56:28:112,714:7:728)	3.517	3.377	3.080	3.061	3.435	3.321	2.989	2.980

- Win(28) and combinations that include it perform very poorly

Best combination?

- Hubicka et al. (2018) recommended AW(28:28:84, 714:7:728)
- We recommend the **WAW(56:28:112, 714:7:728)** scheme
  - Averages long and short windows
  - Good performance across all models and datasets
  - Computationally efficient
- But is the evaluation in terms of MAE sufficient?

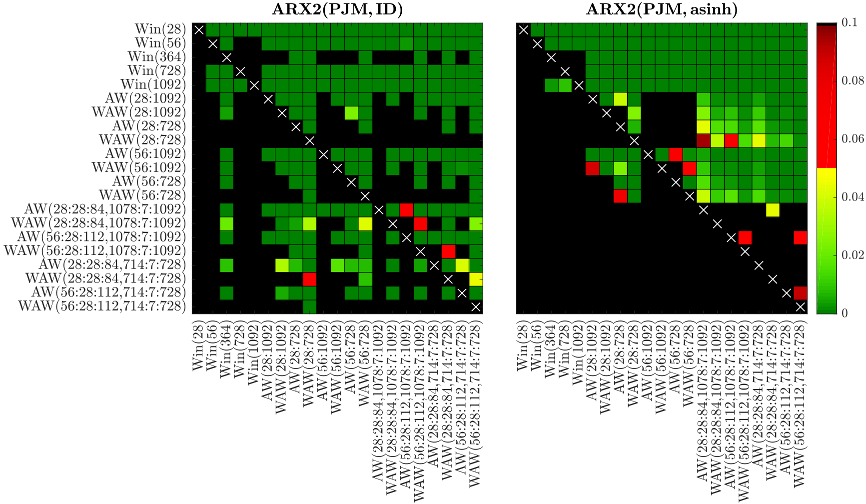
Giacomini-White (2006) test

- For each model pair and each dataset we compute the p-value of the GW test
- Null hypothesis H_0: $\forall_i \phi_i = 0$ in the regression:

$$\Delta_{X,Y,d} = \phi_0 + \phi_1 \Delta_{X,Y,d-1} + \phi_2 \Delta_{X,Y,d-2} + \ldots + \varepsilon_d$$

- Conditional predictive ability
- Better when there is the presence of estimation uncertainty

GW test: p-values, PJM



Conclusions

- Extremely simple yet efficient technique
- Can be applied to different, more complex forecasting techniques
- Brings significant gains in predictions accuracy
- Including longer windows (1092-day) do not increase forecasting accuracy
- None combination is significantly better than **WAW(56:28:112, 714:7:728)**