Project study on ‘A Note on Averaging Day-Ahead Electricity Price Forecasts Across Calibration Windows’

Predictive Analytics Project

Reporter: Yue Wu

Date: 24/03/2023

C O N T E N T

PART 01 - Study Article

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**

I. Introduction

Most day-ahead *electricity price forecasting* (EPF) studies focus on developing model structures that better represent the temporal and inter-variable dependencies, feature (i.e., input variable) selection, implementing faster and more efficient estimation algorithms or finding optimal weights for combined forecasts [1]. However, very few studies in energy forecasting 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.

II. Methodology

Like in many EPF studies, the modeling is implemented here within a 'multivariate' framework [6]. We explicitly use a 'day $\times$ hour', matrix-like structure with $p_{d,h} \equiv \log(P_{d,h})$ representing the electricity log-price for day d and hour h, and consider two models, each consisting of 24 submodels – one for each hour of the day. The **ARX** model is based on a well

PART 01 - Forecast Methodology

01 Dataset

GEFCom2014*(1.1.2011–17.12.2013)*

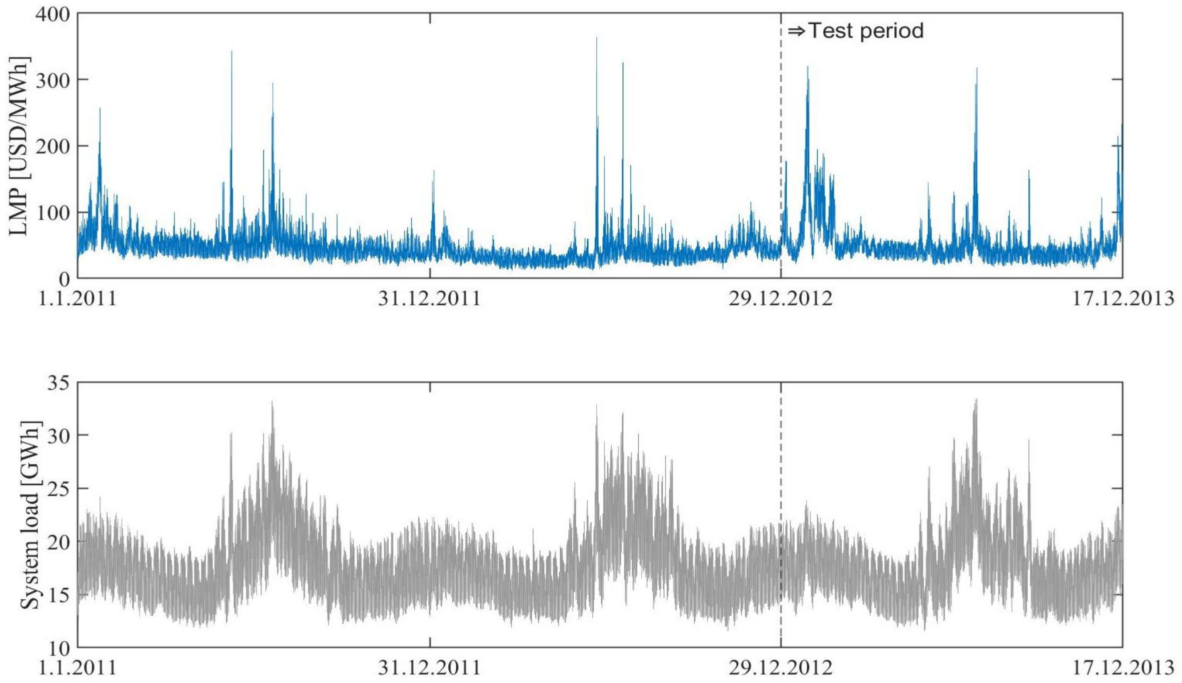


Fig. 1. Hourly locational marginal prices(LMP) and hourly system load of the dataset

02 Calibration window setting

- Calibration window length: 28 - 728 days
- Forecast period: 354 days
- Rolling window scheme

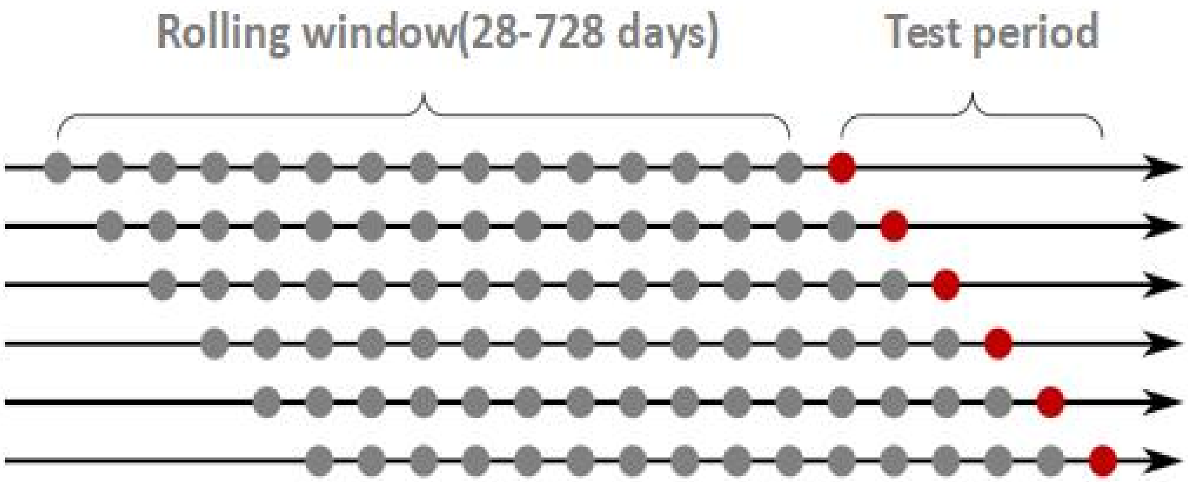


Fig. 2. Rolling window structure

03

Forecast model

- ARX model in a rolling window scheme

Autoregressive structure

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

- ➢ *$p_{d,h}$ - the electricity log-price for day d and hour h*
- ➢ *z_t - the logarithm of the day-ahead load forecast*
- ➢ *D_i - the dummy variable for day-of-the-week i*

03 Forecast model

- NARX model in a rolling window scheme

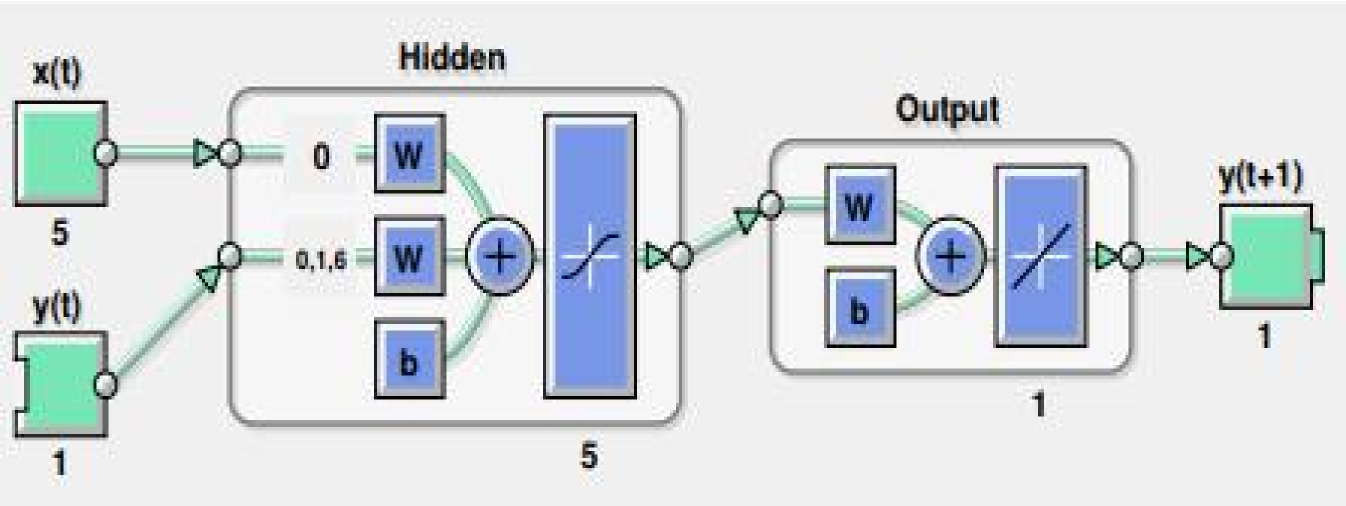


Fig. 3. NARX model structure

04

Forecast evaluation

- Combining various window sets
- Averaging forecasted result
- MAE and Relative changes *(%chng.)*

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

- ➢ $\hat{\varepsilon}_{d,h}$ *- the forecast error for day d and hour h*
- ➢ *D = 354 days*

$$\text{chng.} = \ln(MAE_{Win(728)} / MAE_{AW(\tau)})$$

- ➢ $AW(\tau)$ *- an average of forecast across a set of windows in* τ

- Graphical visualization
- Diebold-Mariano *(DM)* test

PART 02 - Project Workflow

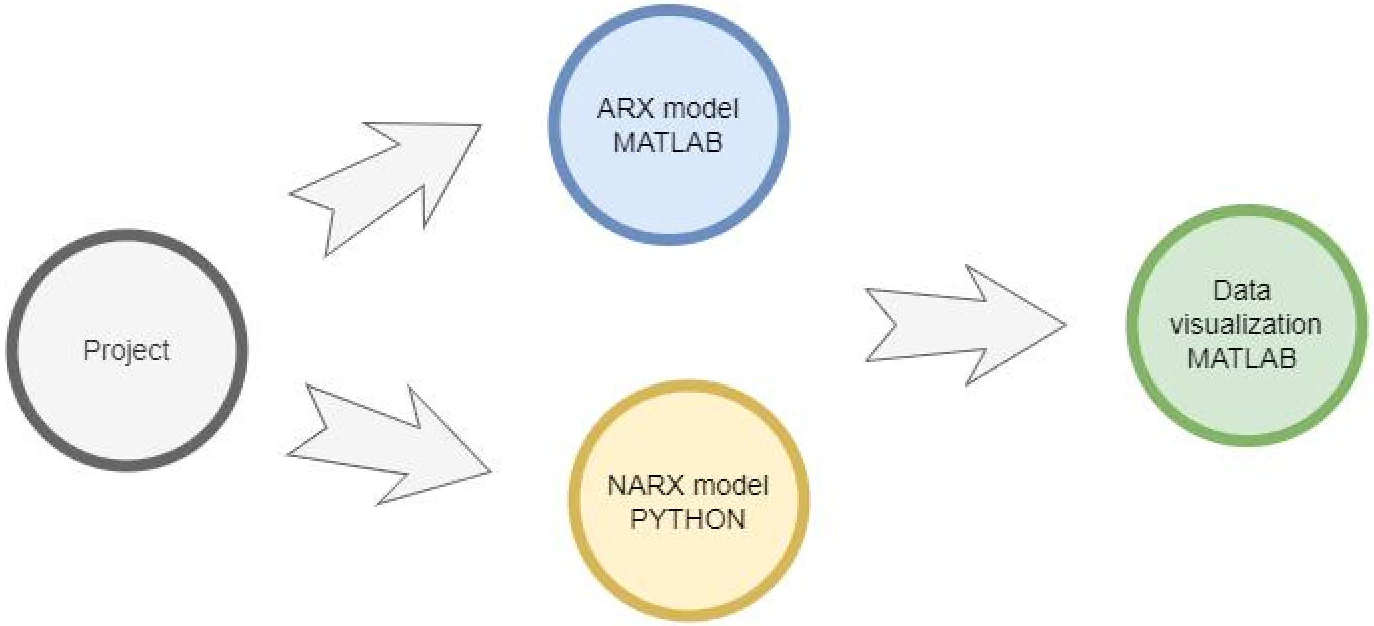


Fig. 4. Project workflow

TABLE 1. MAE and Relative changes (%chng.) for selected calibration window sets

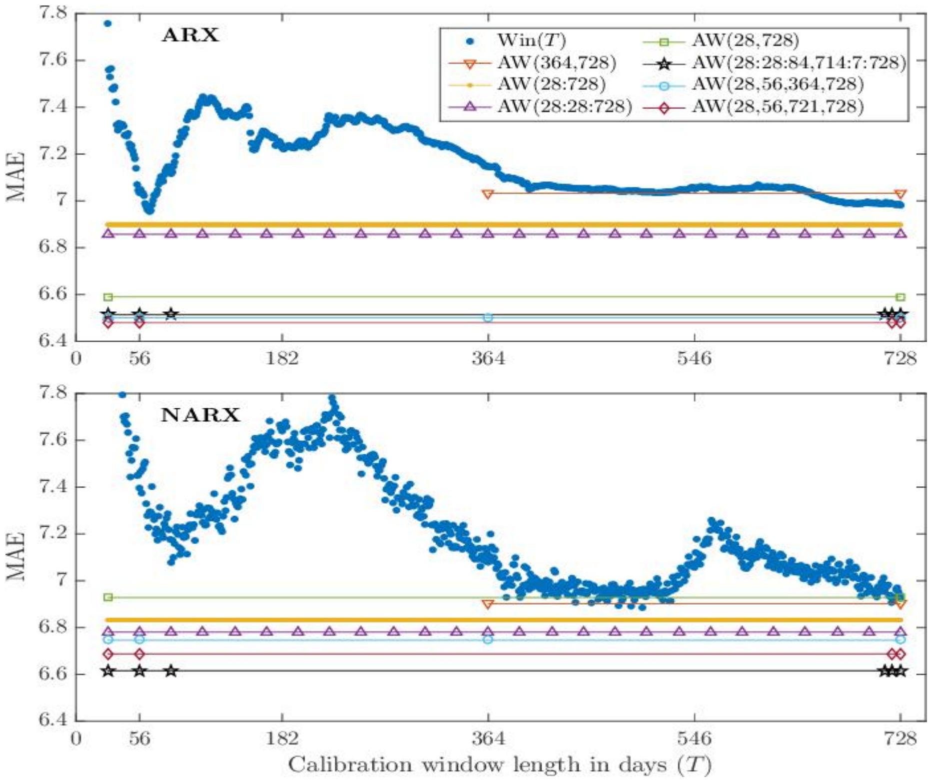
%Chng. compared window

		Article				Project			
		ARX		NARX		ARX		NARX	
Windows		MAE	%chng.	MAE	%chng.	MAE	%chng.	MAE	%chng.
1	Win(28)	7.758	-10.50%	8.394	−19.2%	7.758	-10.50%	10.028	-23.60%
2	Win(364)	7.147	-2.35%	7.079	−2.20%	7.147	-2.35%	8.575	-7.93%
3	AW(364,728)	7.032	-0.72%	6.902	0.34%	7.032	-0.72%	7.811	1.41%
4	Win(728)	6.982	-	9.925	-	6.982	-	7.922	-
5	AW(28:728)	6.898	1.20%	6.832	1.35%	6.898	1.20%	7.071	11.37%
6	AW(28:7:728)	6.891	1.30%	6.793	1.93%	6.891	1.30%	**7.058**	**11.54%**
7	AW(28:14:728)	6.879	1.48%	6.787	2.01%	6.879	1.48%	7.097	10.99%
8	AW(28:28:728)	6.858	1.78%	6.781	2.10%	6.858	1.78%	7.117	10.72%
9	AW(56,728)	6.638	5.05%	6.718	3.03%	6.637	5.05%	7.441	6.26%
10	AW(28,728)	6.591	5.76%	6.928	−0.04%	6.591	5.76%	8.025	-1.30%
11	AW(28:28:84,714:7:728)	6.514	6.93%	**6.616**	**4.57%**	6.514	6.93%	7.132	10.50%
12	AW(28,56,728)	6.509	7.01%	6.888	0.54%	6.509	7.01%	7.48	5.74%
13	AW(28,56,364,728)	6.501	7.13%	6.746	2.61%	6.501	7.13%	7.321	7.89%
14	AW(28,56,721,728)	**6.480**	**7.46%**	6.688	3.49%	**6.48**	**7.46%**	7.229	9.15%

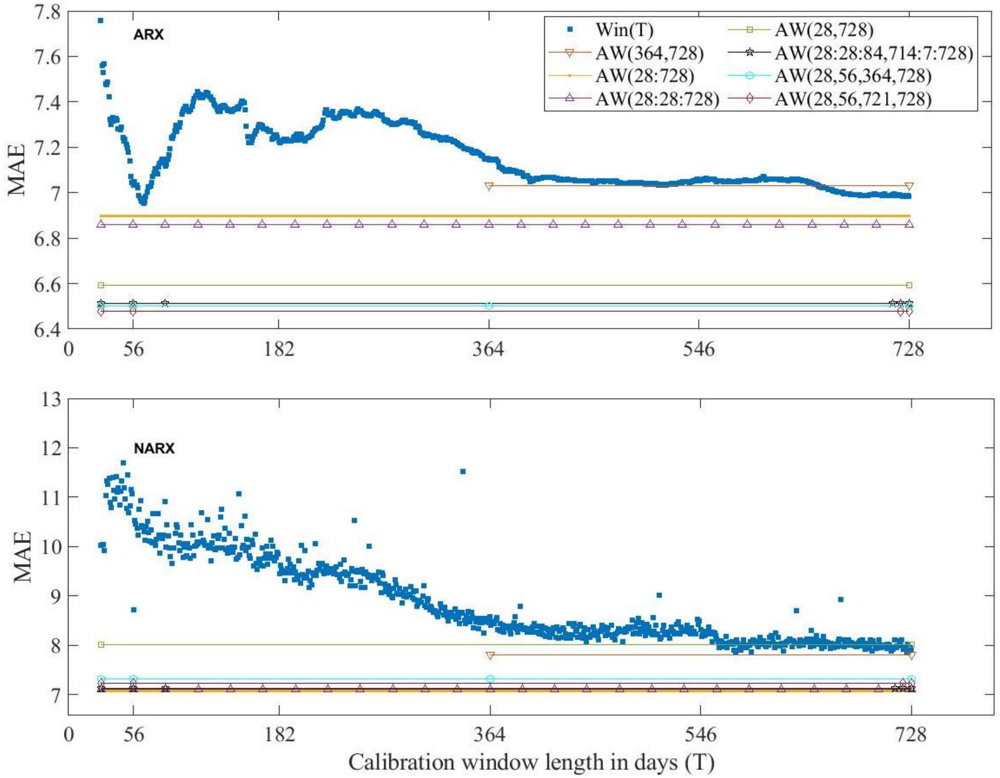
PART 03 - Results comparison & analysis

Fig. 5. MAE of the forecast of single calibration window and various windows sets

Article



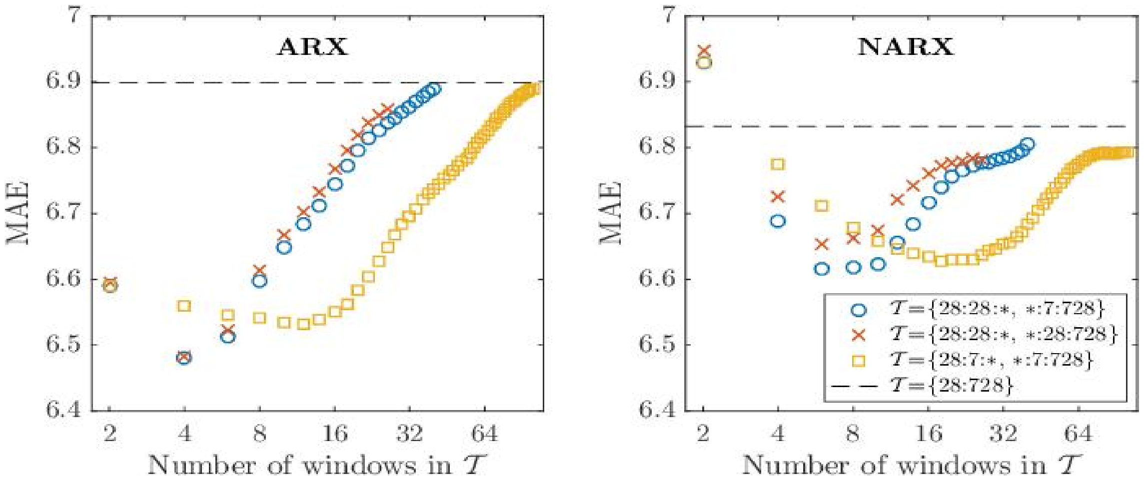
Project



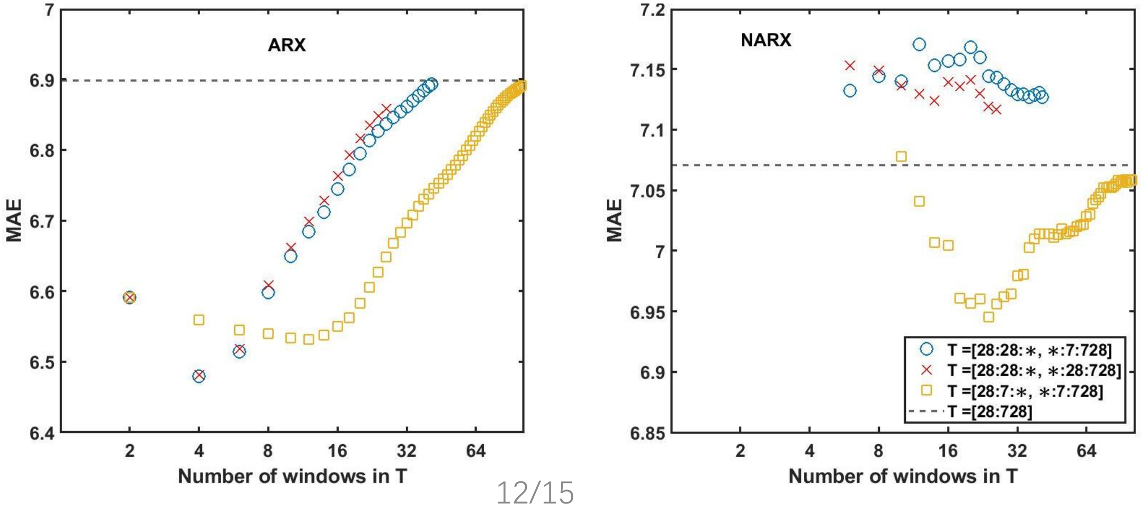
PART 03 - Results comparison & analysis

Fig. 6. MAE of the forecast of iterative windows sets in T = {28:m: , * :n:728} (m,n = 7 or 28)*

Article



Project

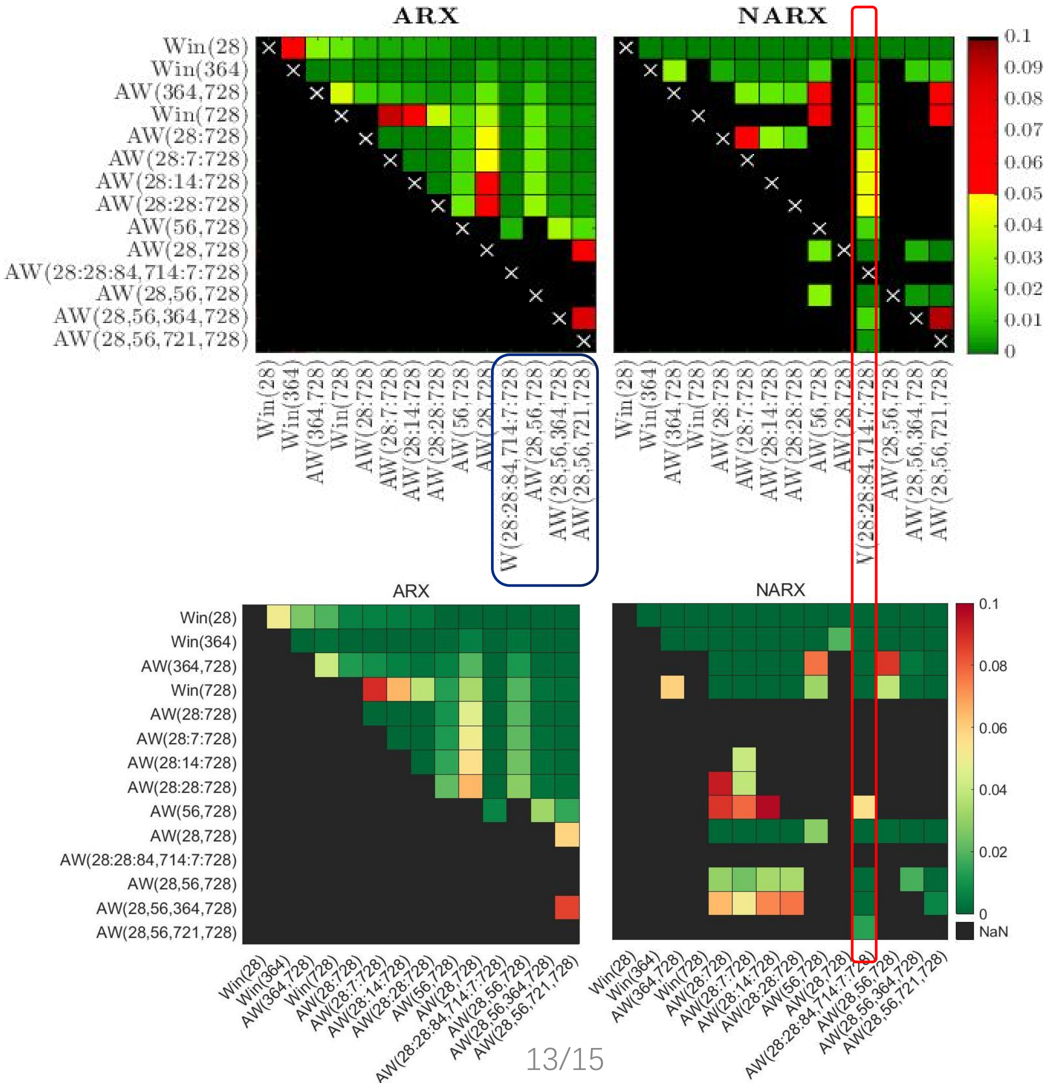


PART 03 - Results comparison & analysis

Fig. 7. Results of the multivariate DM test for selected window sets

Article

Project



PART 03 - Conclusion

- The forecast results of the ARX model is consistent with the original article.
- Due to lack of repeated training, the forecast of NARX model has a large deviation and volatility.
- Combining and averaging forecasts from different windows sets generally yields better results than selecting only one window length.
- More significant accuracy gains can be obtained by averaging over a few windows.

Thank you for listening!