HSC Research Report

HSC/17/01

Variance stabilizing transformations for electricity spot price forecasting

Bartosz Uniejewski[1]
Rafał Weron[2]
Florian Ziel[3]

[1] Faculty of Pure and Applied Mathematics, Wrocław University of Technology, Poland
[2] Department of Operations Research, Wrocław University of Technology, Poland
[3] Faculty of Business Administration and Economics, Universität Duisburg-Essen, Germany

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

Fig. 2. Time series plot of the original (in EUR/MWh; *top*), **asinh$_2$**-transformed (*middle*) and **N-PIT**-transformed EPEX.DE+AT spot prices. The marginal densities are depicted in the right panels.

with inverse:

$$p_{d,h} = \text{sgn}(Y_{d,h}) \left(e^{|Y_{d,h}| - \log c} - \frac{1}{c} \right). \qquad (13)$$

The **mlog**(c) transform is a one parameter family. We use $c = \frac{1}{3}$, which again was selected based on a limited optimization study. Similarly as the standard log-transform and **asinh**, the mirror-log exhibits a log-damping effect, see Fig. 1. Both the **mlog** and **poly** are constructed so that they have a slope of c at the origin. Hence, **mlog**(1) is the same as **boxcox**(0).

The last class of transformations we consider here is based on the so-called *probability integral transform*: $\text{PIT}(X) = F_X(X)$, where F_X is the cumulative distribution function (cdf) of random variable X. The origin of the PIT is not known, but as Gneiting et al. [25] argue, it can be traced back at least to the works of Karl Pearson in the 1930s. In the empirical context we usually do not know the true distribution of X, hence the PIT is rather defined as: $\text{PIT}(X) = \hat{F}_X(X)$, where $\hat{F}_X$ is an estimate (e.g., empirical cdf) or a distributional forecast of F_X. If the latter is perfect, i.e., $\hat{F}_X = F_X$, then $\text{PIT}(X)$ is an independent and uniformly distributed variable. This property can be used to evaluate distributional forecasts [25].

In our context, however, the following transformation will be more useful than the PIT itself:

$$Y_{d,h} = G^{-1}\left(\text{PIT}(P_{d,h})\right) = G^{-1}\left(\hat{F}_{P_{d,h}}(P_{d,h})\right), \qquad (14)$$

where G^{-1} is the inverse of some continuous distribution. Note, that we apply the PIT to original prices, $P_{d,h}$, as this transformation does not require 'normalization' of the inputs. We consider two variants: (i) normal or **N-PIT**, with G^{-1} being the inverse of the standard normal cdf, and (ii) Student-t or **t-PIT**, with G^{-1} being the inverse of the standard Student-t distribution with $\nu = 8$ degrees of freedom (we have also tried other ν's, but lower values yielded too heavy tails, while larger a behavior too similar to that of **N-PIT**).

The **N-PIT** transformation, misleadingly called the Nataf transformation[2], has been used in [27] to 'normalize' Spanish electricity prices. However, the authors have not computed forecasts, only concluded that a 'modified Nataf' transformation (that corrected for the observed zero prices) allowed them to obtain a normal cdf of the transformed prices. The **t-PIT** transformation, on the other hand, has not been used in EPF before. Finally note, that the inverse of Eqn. (14) is given by:

$$P_{d,h} = \hat{F}^{-1}_{P_{d,h}}\left(G(Y_{d,h})\right), \qquad (15)$$

with G being either the standard normal or Student-t cdf.

Since the normal distribution has exponential tails and the Student-t has polynomial, the **N-PIT** has an exponential spike damping behavior and **t-PIT** a polynomial one. Hence, extreme price spikes remain larger for **t-PIT**-transformed data than for **N-PIT**-transformed. In Figure 2 we visualize the effect of applying the **N-PIT** compared to that of original (i.e., untransformed) and **asinh$_2$**-transformed data. We clearly see, that the **N-PIT**-transformed data histogram looks pretty much like a standard normal density. In contrast, **asinh$_2$** preserves the original shape of the histogram of the untransformed data, but damps the price spikes towards the center.

IV. Empirical study

Like Ziel and Weron [13], we use a 730-day (ca. two-year) rolling calibration window testing scheme. First, all considered models are estimated using data from the initial calibration period (i.e., from 31 Jul 2010 to 30 Jul 2012 for the European datasets and from 1 Jan 2011 to 30 Dec 2012 for GEFCom2014) and forecasts for all 24 hours of 31 Jul 2012

[2]The Nataf transformation is a two stage procedure, popularized in reliability engineering by [26]. First, correlated multivariate data is **N-PIT**-transformed independently in each dimension, then the correlation is eliminated by applying the Cholesky factorization.

TABLE II

MEAN ABSOLUTE ERRORS (MAE) ACROSS THE WHOLE TEST PERIOD FOR THE 12 MARKETS (IN COLUMNS) AND THE 16 VARIANCE STABILIZING TRANSFORMATIONS (VSTS; IN ROWS). A HEAT MAP IS USED TO INDICATE BETTER ($\rightarrow$ GREEN) AND WORSE ($\rightarrow$ RED) PERFORMING VSTS. IN THE LAST TWO COLUMNS WE REPORT THE AGGREGATE M.P.D.F.B. ERROR MEASURE, SEE EQN. (18), FOR THE MAE AND THE RMSE.

VST	BELPEX.BE	EPEX.CH	EPEX.DE+AT	EPEX.FR	EXAA.DE+AT	NP.DK1	NP.DK2	NP.SYS	OMIE.ES	OMIE.PT	OTE.CZ	GEFCom2014	m.p.d.f.b.	
													MAE	RMSE
original	6.379	4.019	5.370	5.296	4.282	6.200	5.186	1.890	5.825	5.999	4.670	7.472	5.22%	9.46%
$3\sigma_1$	6.069	3.991	5.180	**4.866**	4.249	5.269	4.948	1.834	6.157	6.326	4.592	8.920	3.98%	6.80%
$3\sigma_2$	6.088	**3.989**	5.199	4.880	4.257	5.379	5.005	1.830	5.858	6.013	4.599	7.720	2.04%	3.04%
$3\sigma\log_1$	6.114	3.993	5.197	4.882	4.270	5.331	5.004	1.829	6.043	6.325	4.605	7.591	2.59%	3.78%
$3\sigma\log_2$	6.137	3.993	5.221	4.905	4.273	5.434	5.055	1.838	5.865	6.055	4.611	7.517	2.29%	1.94%
logistic_1	6.138	4.133	5.369	5.149	4.486	5.296	4.971	1.980	6.817	6.918	4.740	8.042	7.38%	14.29%
logistic_2	**6.047**	4.126	5.250	4.946	4.422	5.303	4.928	1.908	6.578	6.721	4.639	7.921	5.24%	10.81%
asinh_1	6.064	4.074	5.226	4.999	4.289	**5.197**	4.884	1.808	6.134	6.290	4.601	7.022	1.85%	2.73%
asinh_2	6.057	4.080	5.207	4.949	4.288	5.317	4.920	1.803	6.018	6.168	4.590	7.125	1.73%	1.89%
boxcox_1	6.140	4.032	5.231	4.958	4.244	5.372	4.946	1.802	5.845	5.987	4.589	7.074	1.28%	1.08%
boxcox_2	6.153	4.035	5.242	4.967	4.255	5.523	4.988	1.808	5.842	5.989	4.596	7.152	1.80%	1.62%
poly_1	6.113	4.016	5.211	4.923	4.234	5.284	4.939	1.798	5.854	6.001	4.579	7.066	0.93%	**0.60%**
poly_2	6.134	4.018	5.226	4.933	4.245	5.457	4.987	1.807	5.841	5.993	4.587	7.173	1.54%	0.99%
mlog_1	6.098	4.020	5.206	4.924	4.235	5.257	4.928	**1.796**	5.870	6.016	4.577	7.049	0.86%	**0.60%**
mlog_2	6.118	4.023	5.220	4.928	4.246	5.419	4.976	1.803	5.850	6.001	4.584	7.158	1.42%	0.86%
N-PIT	6.054	4.004	**5.162**	4.890	**4.188**	5.205	**4.847**	1.826	5.854	5.993	**4.553**	7.129	**0.45%**	1.47%
t-PIT	6.154	4.089	5.197	4.961	4.264	5.264	4.901	1.868	**5.824**	**5.984**	4.611	**6.991**	1.37%	1.13%

(respectively, 31 Dec 2012) are determined. Then the window is rolled forward by one day, the models are reestimated and forecasts for all 24 hours of the next day are computed. This procedure is repeated until the predictions for the 24 hours of 28 Jul 2016 (respectively, 17 Dec 2013) are computed. Note, that for the European datasets the out-of-sample test period covers roughly four years (1459 days) of data and for the GEFCom2014 dataset with only 352 days.

A. *MAE, RMSE and m.p.d.f.b error measures*

As the main evaluation criterion we consider the *Mean Absolute Error* (MAE) for the full out-of-sample test period of $D = 1459$ days (for GEFCom2014 only 352 days). It is computed for each VST and dataset as:

$$\text{MAE} = \frac{1}{24D} \sum_{d=1}^{D} \sum_{h=1}^{24} |\widehat{\varepsilon}_{d,h}|, \qquad (16)$$

where $\widehat{\varepsilon}_{d,h}$ denotes the estimated forecasting error for day d and hour h. The MAE errors are reported for the 16 considered VSTs and all 12 datasets in Table II. We have also analyzed *Root Mean Square Errors* (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{24D} \sum_{d=1}^{D} \sum_{h=1}^{24} \widehat{\varepsilon}_{d,h}^{2}}, \qquad (17)$$

but the results were similar and hence are not reported nor analyzed here due to space limitations (but are available from the authors upon request). Only in Table II we provide for comparison an aggregate measure of fit based on the RMSE – the m.p.d.f.b. – as defined in (18) below. Although there are some changes in the ranking, the overall picture is very similar.

Given the large number of results it is hard to rank the VSTs. To tackle this issue, following [13], we introduce the *mean percentage deviation from the best* (m.p.d.f.b.) VST, which is inspired by the m.d.f.b. measure used in [11], [28] for comparing models. The m.p.d.f.b. measure for VST i indicates how similar is this VST's performance to the 'optimal VST' composed of the best performing VST for each of the 12 datasets:

$$\text{m.p.d.f.b.}_i = \frac{1}{12} \sum_{j=1}^{12} \frac{|\text{ERR}_{i,j} - \text{ERR}_{\text{best VST},j}|}{\text{ERR}_{\text{best VST},j}} \times 100\%, \quad (18)$$

where $\text{ERR}_{\text{best VST},j} = \min_{1 \le i \le 17} \text{ERR}_{i,j}$ and ERR can be the MAE, the RMSE or any other error measure for point forecasts. The m.p.d.f.b. measure is reported for the original data and 16 considered VSTs in the last two columns of Table II.

From Table II we can see the dominance of the **N-PIT** over the competitors in terms of MAE. It has the lowest m.p.d.f.b., nearly twice smaller than the next best transform, i.e., the $\mathbf{mlog}_1$. Also for four datasets (EPEX.DE+AT, EXAA.DE+AT, NP.DK2 and OTE.CZ) it is the best performer and second best for another two (BELPEX.BE and NP.DK1). However, for some markets (NP.SYS, OMIE.ES and GEFCom2014) it does not excel. The **t-PIT** transformation is the best for three markets (OMIE.ES, OMIE.PT and GEFCom2014), but the overall performance does not seem to be robust and the **t-PIT** forecasting accuracy is poor for some markets (especially EPEX.CH and NP.SYS). The $\mathbf{3\sigma}$, $\mathbf{3\sigma log}$ and **asinh** transformations show moderate MAE forecasting performance. Still, their m.p.d.f.b. is better than that of the original (i.e, untransformed) data and sometimes even that of the best transformation for a given market ($\mathbf{3\sigma}_1$ for EPEX.CH, $\mathbf{3\sigma}_2$ for EPEX.FR and $\mathbf{asinh}_1$ for NP.DK1).

In terms of RMSE, both **N-PIT** and **t-PIT** are on average outperformed by the generalizations of the Box-Cox transform, i.e., **poly** and **mlog**, which show in general a very similar behavior across all 12 datasets. The **boxcox** and **PIT** transformations follow closely, while the remaining ones lag behind. Interestingly, the **t-PIT** has a better RMSE forecasting accuracy than the **N-PIT**. Furthermore, we see that the **logistic** transformations are the only ones which are worse in terms of m.p.d.f.b. (both MAE and RMSE-based) than the original (i.e., untransformed) data. Still, somewhat surprisingly **logistic**$_2$ is the best choice in terms of MAE for the Belgian market, which nicely illustrates that the overall behavior can depend on the considered market and data structure. Finally, we want to remark that there is no clear tendency if a scaling by set_1 (median, MAD) or set_2 (mean, std) is preferable, although for the Box-Cox type transformations (**boxcox**, **poly** and **mlog**) set_1 leads to marginally better predictions in terms of m.p.d.f.b., see Table II.

B. Diebold-Mariano tests

The MAE values analyzed in Section IV-A can be used to provide a ranking of transformations, but not statistically significant conclusions on the outperformance of the forecasts of one transformation by those of another. Therefore, we also computed the Diebold-Mariano (DM) test [14], which takes the correlation structure into account. It tests forecasts of each pair of transformations against each other.

In the EPF literature, the DM test is usually performed separately for each of the 24 hours of the day [1]. However, Ziel and Weron [13] recently introduced a different approach, where only one statistic for each pair of models (here: VSTs) is computed based on the 24-dimensional vector of errors for each day, and called it the *multivariate* or *vectorized* DM test. Following [13], denote by $\widehat{\boldsymbol{\varepsilon}}_{X,d} = (\widehat{\varepsilon}_{X,d,1}, \dots, \widehat{\varepsilon}_{X,d,24})'$ and $\widehat{\boldsymbol{\varepsilon}}_{Y,d} = (\widehat{\varepsilon}_{Y,d,1}, \dots, \widehat{\varepsilon}_{Y,d,24})'$ the vectors of out-of-sample errors for day d of VSTs X and Y, respectively. Then the multivariate loss differential series:

$$\Delta_{X,Y,d} = \|\widehat{\boldsymbol{\varepsilon}}_{X,d}\|_p - \|\widehat{\boldsymbol{\varepsilon}}_{Y,d}\|_p, \qquad (19)$$

defines the differences of errors in the $\|\cdot\|_p$-norm, i.e., $\|\widehat{\boldsymbol{\varepsilon}}_{X,d}\|_p = (\sum_{h=1}^{24} |\widehat{\varepsilon}_{X,d,h}|^p)^{1/p}$ for $p = 1, 2$. For each model pair and each dataset we compute the p-value of two one-sided DM tests: (i) a test with the null hypothesis $H_0 : E(\Delta_{X,Y,d}) \leq 0$, i.e., the outperformance of the forecasts of Y by those of X, and (ii) the complementary test with the reverse null $H_0^R : E(\Delta_{X,Y,d}) \geq 0$, i.e., the outperformance of the forecasts of X by those of Y. As in the standard DM test, we assume that the loss differential series is covariance stationary.

To jointly evaluate the performance of the VSTs across all 11 European markets we introduce yet another variant of the DM test, in which the norm is computed not only for all hours but all datasets as well. Note, that we exclude GEFCom2014 from this analysis due to a much shorter test period. The computations are analogous, only this time the multivariate loss differential series across all 11 European markets, i.e., M_1 = BELPEX.BE, ..., M_{11} = OTE.CZ, is given by:

$$\Delta_{X,Y,d,M} = \|\widehat{\boldsymbol{\varepsilon}}_{X,d}\|_p - \|\widehat{\boldsymbol{\varepsilon}}_{Y,d}\|_p, \qquad (20)$$

where now $\|\widehat{\boldsymbol{\varepsilon}}_{X,d}\|_p = (\sum_{j=1}^{11} \sum_{h=1}^{24} |\widehat{\varepsilon}_{X,d,h,M_j}|^p)^{1/p}$ and $\widehat{\boldsymbol{\varepsilon}}_{X,d} = (\widehat{\varepsilon}_{X,d,1,M_1}, \dots, \widehat{\varepsilon}_{X,d,24,M_{11}})'$ is the vector of out-of-sample errors all hours and all markets on day d.

In Figure 3 we plot the results for the multivariate DM-test for each market using the $\|\cdot\|_1$-norm, i.e., for $p = 1$ in Eqn. (19), while in Figure 4 the aggregated DM-test as defined in Eqn. (20). In both figures we see the corresponding p-values of the conducted pairwise comparisons. Green and yellow squares indicate statistical significance at the 5% level. For instance, we see in Fig. 3 for BELPEX.BE that the first row is completely green. So every transformation significantly improved the forecasting accuracy compared to the original untransformed prices. Similarly for EXAA.DE+AT, the column which corresponds to the **N-PIT** is dark green, meaning that **N-PIT** leads to significantly better forecasts than all other transformations under consideration.

Regarding the aggregated DM-tests in Fig. 4, we see that all transformations lead to significantly better predictions than the original data, except for **logistic**. This result holds for the $\|\cdot\|_1$- and the $\|\cdot\|_2$-norm, even though the latter tends to return higher p-values. For the $\|\cdot\|_1$-norm the **N-PIT** transformation leads to significantly better forecasts than all other options, which emphasizes the findings from Table II. But for the $\|\cdot\|_2$-norm the results are not that clear-cut. Still **poly**$_1$ and **mlog**$_1$ lead to forecasts that outperform most of the competitors, except each other, $\boldsymbol{3\sigma_2}$ and **N-PIT**. Heaving this in mind, the **N-PIT** seems to be the overall best performer. It leads to significantly better predictions than all other transformations within the robust evaluation framework with respect to the $\|\cdot\|_1$-norm and not significantly worse than the best transformations in the $\|\cdot\|_2$-norm framework. However, if the evaluation focus is on spike detection then we suggest to use the **mlog**$_1$ as it has a very similar performance to the **poly**$_1$, but requires only one shape parameter (i.e., c).

V. Conclusions

We have conducted an extensive EPF study across 12 major markets to evaluate different variance stabilizing transformations. In line with the guidelines set forth in [1], we have used two variants of the Diebold-Mariano (DM) test to formally assess the statistical significance of the forecasting performance. The obtained results suggest that the choice of the optimal transformation depends on the forecasting framework and the considered dataset.

However, while for individual markets specifically tailored transformations can yield better results, due to the increasing demand for joint modeling of multiple markets, robust cross-market transformations may turn out to be very useful. In particular, the *probability integral transform*-based **N-PIT** yields very robust results and promising forecasting accuracy in terms of MAE. It leads to forecasts that significantly outperform all other competitors across all markets, according to the $\|\cdot\|_1$-norm based DM test. However, if the forecasting focus is

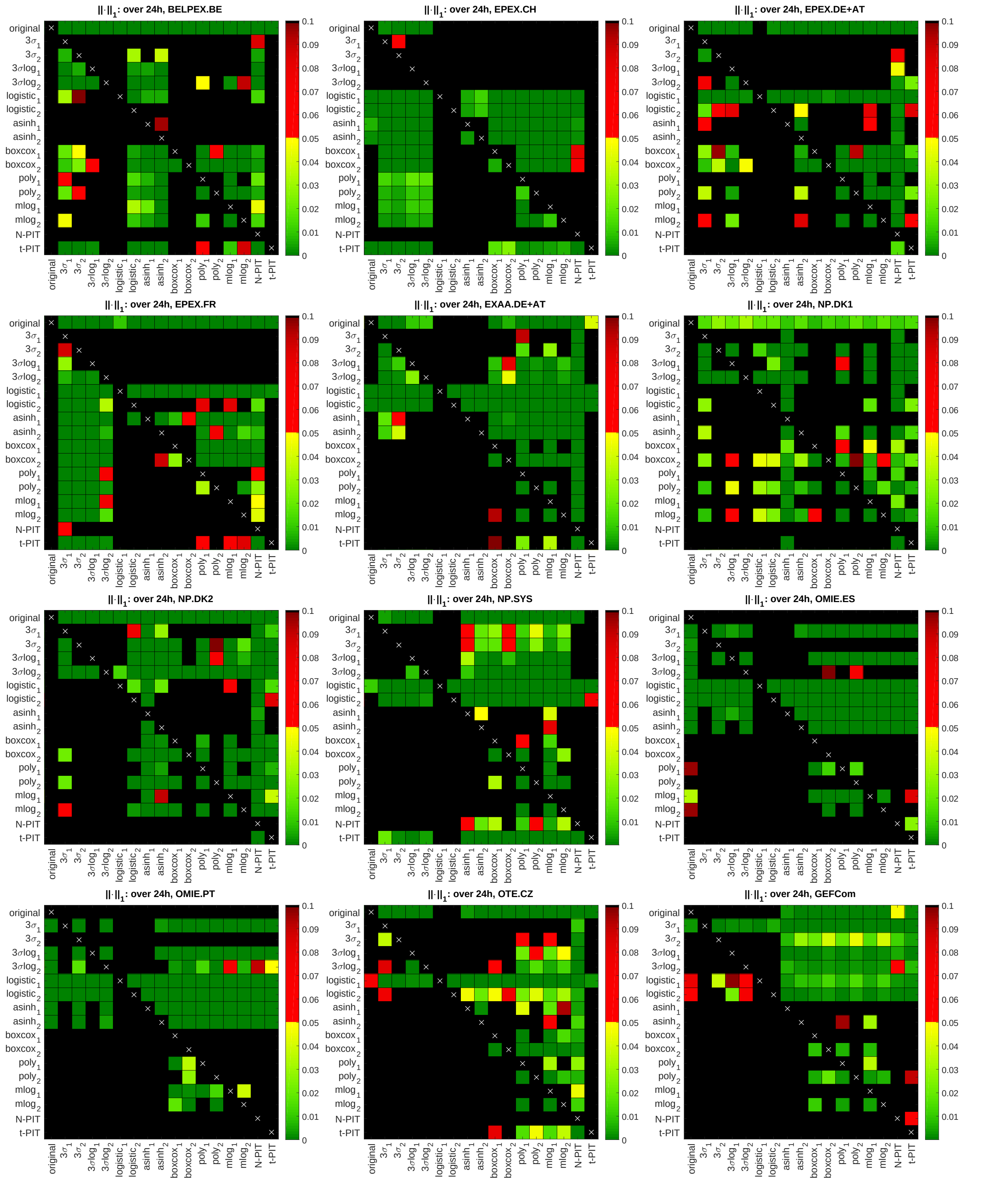

Fig. 3. Results of the 'multivariate' DM test defined by the multivariate loss differential series in Eqn. (19) with $p = 1$, i.e., in the $\| \cdot \|_1$-norm, for all 12 datasets. Like in Figure 4, we use a heat map to indicate the range of the p-values – the closer they are to zero ($\rightarrow$ dark green) the more significant is the difference between the forecasts of a model on the X-axis (better) and the forecasts of a model on the Y-axis (worse).

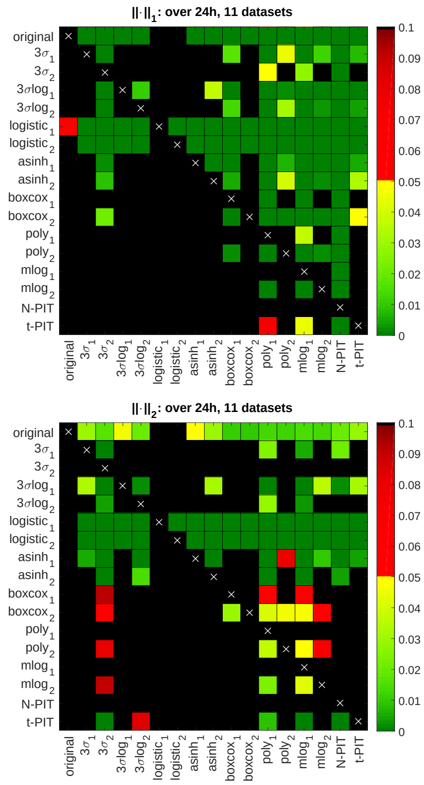

Fig. 4. Results of the 'multivariate' DM test defined by the multivariate loss differential series in Eqn. (20) with $p = 1$ (*top*) and $p = 2$ (*bottom*) across all 11 European datasets. We use a heat map to indicate the range of the p-values – the closer they are to zero ($\rightarrow$ dark green) the more significant is the difference between the forecasts of a model on the X-axis (better) and the forecasts of a model on the Y-axis (worse).

on spike sensitive measures (like the RMSE), then the newly introduced **poly** and **mlog** transforms tend to perform better. Given that the **mlog** requires only one shape parameter (c), while **poly** needs two (λ, c), we suggest to use the former in such a context.

Our study can be further expanded in several directions. In particular, we report results for only one expert, regression-based model. Although we have also considered several other expert and autoregressive models from [13], [15] and the results were qualitatively the same, we can only conjecture that our conclusions will hold for more complex models, like parameter rich structures estimated via the LASSO [13], [15], [19] or well performing neural network feature selection algorithms [29]. Moreover, in this study we have restricted ourselves to symmetric transformations. Future research could elaborate on asymmetric functions, which may yield an even better forecasting performance.

References

[1] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.

[2] E. Fanone, A. Gamba, and M. Prokopczuk, "The case of negative day-ahead electricity prices," *Energy Economics*, vol. 35, pp. 22–34, 2013.

[3] J. Janczura, S. Trück, R. Weron, and R. Wolff, "Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling," *Energy Economics*, vol. 38, pp. 96–110, 2013.

[4] P. J. Huber and E. M. Ronchetti, *Robust Statistics, 2nd ed.* Wiley, 2009.

[5] J. H. Zhao, Z. Y. Dong, X. Li, and K. P. Wong, "A framework for electricity price spike analysis with advanced data mining methods," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 376–385, 2007.

[6] R. Weron, "Heavy-tails and regime-switching in electricity prices," *Mathematical Methods of Operations Research*, vol. 69, no. 3, pp. 457–473, 2009.

[7] D. Chen and D. W. Bunn, "Analysis of the nonlinear response of electricity prices to fundamental and strategic factors," *IEEE Transactions on Power Systems*, vol. 25, p. 595–606, 2010.

[8] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.

[9] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. Wiley, 2002.

[10] J. Contreras, R. Espínola, F. Nogales, and A. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, 2003.

[11] R. Weron and A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *International Journal of Forecasting*, vol. 24, pp. 744–763, 2008.

[12] S. Schneider, "Power spot price models with negative prices," *Journal of Energy Markets*, vol. 4, no. 4, pp. 77–102, 2011.

[13] F. Ziel and R. Weron, "Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate models," *Energy Economics*, 2017, submitted.

[14] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 13, pp. 253–263, 1995.

[15] B. Uniejewski, J. Nowotarski, and R. Weron, "Automated variable selection and shrinkage for day-ahead electricity price forecasting," *Energies*, vol. 9, no. 8, p. 621, 2016.

[16] M. Burger, B. Graeber, and G. Schindlmayr, *Managing energy risk: An integrated view on power and other energy markets*. Wiley, 2007.

[17] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. John Wiley & Sons, Chichester, 2006.

[18] N. Karakatsani and D. W. Bunn, "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients," *International Journal of Forecasting*, vol. 24, pp. 764–785, 2008.

[19] F. Ziel, "Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4977–4987, 2016.

[20] J. S. Armstrong, Ed., *Principles of Forecasting: A handbook for researchers and practitioners*. Kluver, 2001.

[21] N. Hamdy, *Applied Signal Processing: Concepts, Circuits, and Systems*. CRC Press, Boca Raton, 2008.

[22] F. Günther and S. Fritsch, "neuralnet: Training of neural networks," *The R journal*, vol. 2, no. 1, pp. 30–38, 2010.

[23] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and practice*. Online at http://otexts.org/fpp/, 2013.

[24] R. Sakia, "The Box-Cox transformation technique: A review," *The Statistician*, vol. 41, no. 2, pp. 169–178, 1992.

[25] T. Gneiting, F. Balabdaoui, and A. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society B*, vol. 69, pp. 243–268, 2007.

[26] A. Der Kiureghian and P.-L. Liu, "Structural reliability under incomplete probability information," *Journal of Engineering Mechanics*, vol. 112, no. 1, pp. 85–104, 1986.

[27] G. Diaz and E. Planas, "A note on the normalization of Spanish electricity spot prices," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2499–2500, 2016.

[28] J. Nowotarski, E. Raviv, S. Trück, and R. Weron, "An empirical comparison of alternate schemes for combining electricity spot price forecasts," *Energy Economics*, vol. 46, pp. 395–412, 2014.

[29] O. Abedinia, N. Amjady, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62–74, 2017.

HSC Research Report Series 2017

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Variance stabilizing transformations for electricity spot price forecasting* by Bartosz Uniejewski, Rafał Weron and Florian Ziel