Variance stabilizing transformations for electricity spot price forecasting: Simple autoregressive models and beyond*

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

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

*Partially supported by the National Science Center (NCN, PL) through grant no. 2015/17/B/HS4/00334 energies

Energies **2016**, *9*, 621; doi:10.3390/en9080621 MDPI

Article OPEN ACCESS

Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting

Bartosz Uniejewski, Jakub Nowotarski and Rafał Weron *

Department of Operations Research, Wrocław University of Technology, 50-370 Wrocław, Poland;
uniejewskibartosz@gmail.com
* Correspondence: rafal.we

Academic Editor: Javier Cont
Received: 5 July 2016; Accepte

Abstract: In day-ahead e
Conducting an empirical st
datasets from three major
procedures (single-step eli
we show that using the
commonly-used EPF mode
in EPF before, stands out a

IEEE Xplore® Digital Library

IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. XX, NO. X, 2017/2018

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 (VSTs) 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* (PIT) combined with the standard Gaussian distribution yields the best approach, significantly better than many of the considered alternatives.

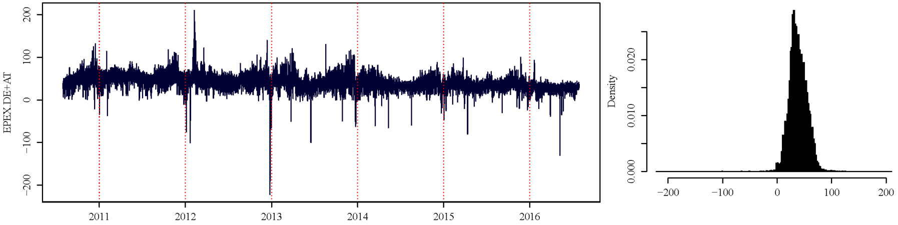
$\hat{F}_Z$	Estimate or distributional forecast of F_Z
$\widehat{\varepsilon}_{X,d}$	Vector of 24 hourly out-of-sample errors for day d of VST X, i.e., $(\widehat{\varepsilon}_{X,d,1},\ldots,\widehat{\varepsilon}_{X,d,24})'$
$\|\cdot\|_p$	p-norm, i.e., $\|\widehat{\varepsilon}_{X,d}\|_p = (\sum_{h=1}^{24}\|\widehat{\varepsilon}_{X,d,h}\|^p)^{1/p}$
$\Delta_{X,Y,d}$	Loss differential series of the multivariate Diebold-Mariano (DM) test, see Eqn. (18)
$\Delta_{X,Y,d,M}$	Loss differential series of the multivariate DM test across 11 European markets (M_1,...,M_{11}), see Eqn. (19)

I. Introduction

Over the last two decades *short-term* (also called *spot* or *day-ahead*; for a discussion see [1]) *electricity price forecasting* (EPF) has joined short-term load forecasting as one

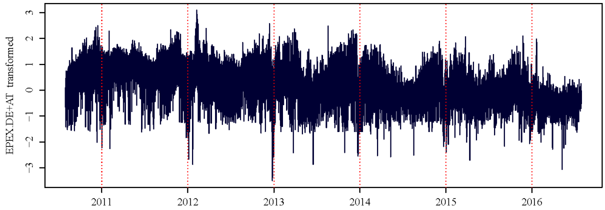
The problem

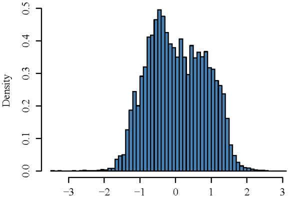
- Most electricity spot price series exhibit price spikes
- These 'outliers' may impact the obtained model estimates



The solution

Transform prices to reduce spike severity and stabilize the variance:





- The logarithm is the most commonly used transform
- ... but for datasets with very close to zero ($\rightarrow$ Spanish) or negative ($\rightarrow$ German) prices the log-transform is not feasible
- **What reasonable choices do we have then?**

Study setup

Uniejewski, Weron & Ziel (2017/18, TPWRS)

A comprehensive study involving **12** datasets:

Electricity market & region	Source	Mean	Std	Median	MAD	Min	Max
6-year period (30 Jul 2010 to 28 Jul 2016, 2189×24 hourly observations)							
BELPEX, Belgium	belpex.be	44.49	22.15	44.94	19.54	−200.00	2999.00
EPEX, Switzerland	epexspot.com	44.60	17.68	43.98	20.91	−45.68	300.04
EPEX, Germany & Austria	epexspot.com	38.47	16.71	37.31	19.01	−221.99	210.00
EPEX, France	epexspot.com	41.61	22.04	41.50	20.62	−200.00	1938.50
EXAA, Germany & Austria	exaa.at	38.70	15.65	37.37	18.56	−50.92	175.74
Nord Pool, West Denmark	nordpoolspot.com	35.31	24.23	33.51	17.93	−200.00	2000.00
Nord Pool, East Denmark	nordpoolspot.com	37.23	21.53	34.65	18.96	−200.00	2000.00
Nord Pool, System price	nordpoolspot.com	34.07	15.12	31.95	16.26	1.14	224.97
OMIE, Spain	omie.es	45.23	15.72	47.94	17.56	0.00	112.00
OMIE, Portugal	omie.es	45.11	15.97	47.96	17.81	0.00	145.00
OTE, Czechia	ote-cr.cz	38.42	16.16	37.38	18.74	−150.00	170.00
3-year period (1 Jan 2011 to 17 Dec 2013, 1082×24 hourly observations)							
GEFCom2014 competition	Hong et al. (2016)	48.19	26.18	42.87	23.97	12.52	363.80

... **16** variance stabilizing transformations (VSTs) and **2** model classes

Prior to applying VSTs

We 'normalize' the spot prices:

$$p_{d,h} = \frac{P_{d,h} - a}{b}$$

and consider 2 sets of 'normalizing' parameters:

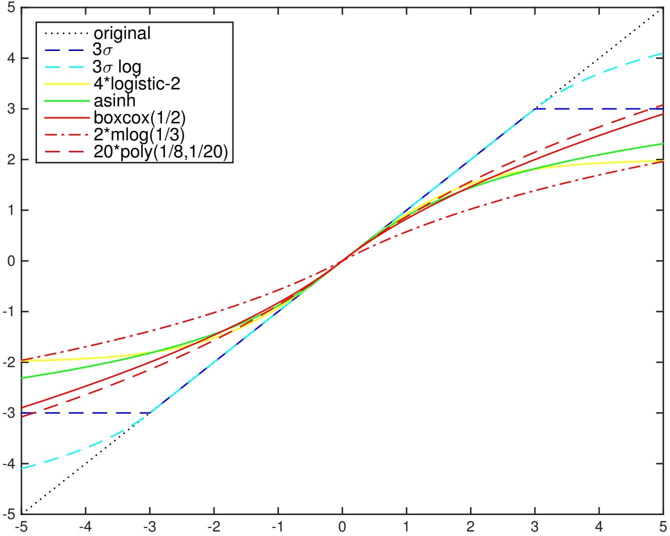
set_1: (a,b) = (median,MAD)

- a is the median of $P_{d,h}$ and b is the sample *median absolute deviation* (MAD)

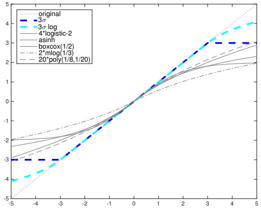
set_2: (a,b) = (mean,std)

- a is the mean and b is the standard deviation of $P_{d,h}$

Variance Stabilizing Transformations (VSTs)



3σ-type transformations



- **3$\boldsymbol{\sigma}$** clipping/winsorizing ('classical')

$$Y_{d,h} = \begin{cases} 3\ \mathrm{sgn}(p_{d,h}) & \text{for } |p_{d,h}| > 3, \\ p_{d,h} & \text{for } |p_{d,h}| \leqslant 3, \end{cases}$$

  - inverse

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

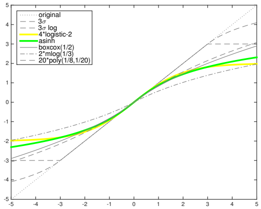
- **3$\boldsymbol{\sigma}$log** damping (Shahidehpour et al., 2002; Weron, 2006)

$$Y_{d,h} = \begin{cases} \mathrm{sgn}(p_{d,h})\{\log(|p_{d,h}| - 2) + 3\} & \text{for } |p_{d,h}| > 3, \\ p_{d,h} & \text{for } |p_{d,h}| \leqslant 3, \end{cases}$$

  - inverse

$$p_{d,h} = \begin{cases} \mathrm{sgn}(Y_{d,h})\left(e^{|Y_{d,h}|-3} + 2\right) & \text{for } |Y_{d,h}| > 3, \\ Y_{d,h} & \text{for } |Y_{d,h}| \leqslant 3. \end{cases}$$

Two special functions



- **logistic** (link function)

$$Y_{d,h} = \frac{1}{1 + e^{-p_{d,h}}}$$

  - inverse

$$p_{d,h} = \log\left(\frac{Y_{d,h}}{1 - Y_{d,h}}\right).$$

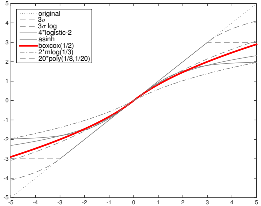
- area hyperbolic sine or **asinh** (Schneider, 2011, JEM)

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

  - inverse

$$p_{d,h} = \sinh(Y_{d,h})$$

Box-Cox transformation



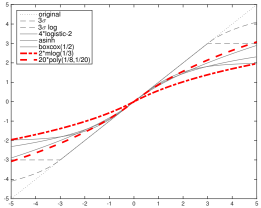
- **boxcox** (‘corrected’ for negative values)

$$Y_{d,h} = \mathrm{sgn}(p_{d,h}) \begin{cases} \frac{\left(|p_{d,h}|+1\right)^{\lambda}-1}{\lambda} & \text{for } \lambda > 0, \\ \log(|p_{d,h}|+1) & \text{for } \lambda = 0, \end{cases}$$

  - inverse

$$p_{d,h} = \mathrm{sgn}(Y_{d,h}) \begin{cases} (\lambda|Y_{d,h}|+1)^{\frac{1}{\lambda}} - 1 & \text{for } \lambda > 0, \\ e^{|Y_{d,h}|} - 1 & \text{for } \lambda = 0. \end{cases}$$

New Box-Cox generalizations



- **poly** $(\lambda > 0)$

$$Y_{d,h} = \operatorname{sgn}(p_{d,h}) \left[\left| |p_{d,h}| + \left(\frac{c}{\lambda}\right)^{\frac{1}{\lambda-1}} \right|^{\lambda} - \left(\frac{c}{\lambda}\right)^{\frac{\lambda}{\lambda-1}} \right]$$

  - inverse

$$p_{d,h} = \operatorname{sgn}(Y_{d,h}) \left[\left(|Y_{d,h}| + \left(\tfrac{c}{\lambda}\right)^{\frac{\lambda}{\lambda-1}} \right)^{\frac{1}{\lambda}} - \left(\tfrac{c}{\lambda}\right)^{\frac{1}{\lambda-1}} \right].$$

- mirror-log or **mlog** $(\lambda{=}0)$

$$Y_{d,h} = \operatorname{sgn}(p_{d,h}) \left[\log\left(|p_{d,h}| + \frac{1}{c} \right) + \log(c) \right]$$

  - inverse $\quad p_{d,h} = \operatorname{sgn}(Y_{d,h}) \left[e^{|Y_{d,h}| - \log c} - \frac{1}{c} \right]$

Probability Integral Transform (PIT)

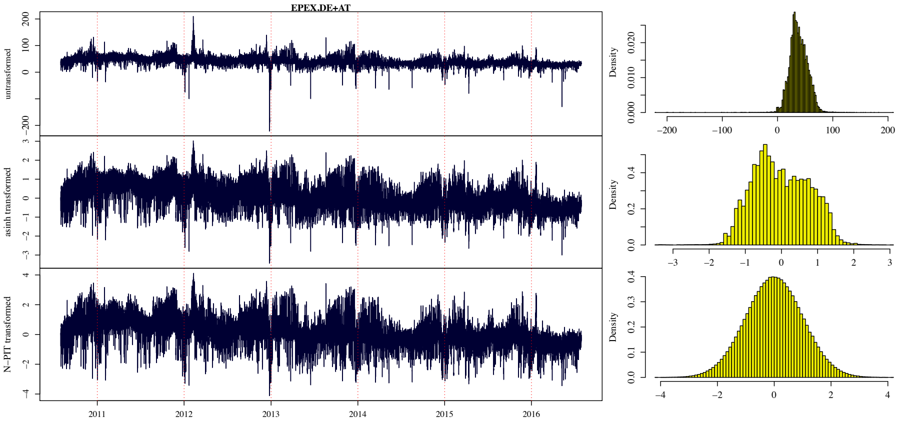
$$Y_{d,h} = G^{-1}\left(\mathrm{PIT}(P_{d,h})\right) = G^{-1}\left(\hat{F}_{P_{d,h}}(P_{d,h})\right)$$

where:

- G^{-1} is the inverse of some continuous distribution
- PIT is an estimate (i.e., empirical cdf) or a distributional forecast of F_P
- We consider two variants:
  - normal or **N-PIT**
  - Student-t or **t-PIT**

Sample results

(Original, **asinh**- and **N-PIT**-transformed EPEX.DE+AT price series)



Model classes

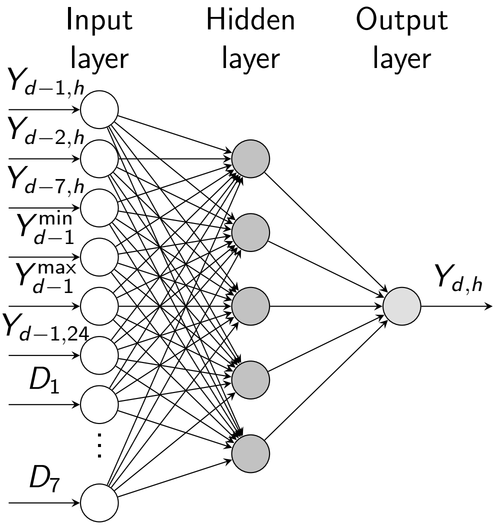
Benchmark #1: Autoregressive 'expert' model

A well performing multivariate 'expert' model (Ziel & Weron, 2017):

$$
\begin{aligned}
Y_{d,h} = & \beta_{h,1} + \underbrace{\beta_{h,2} Y_{d-1,h} + \beta_{h,3} Y_{d-2,h} + \beta_{h,4} Y_{d-7,h}}_{\text{autoregressive terms}} \\
& + \underbrace{\beta_{h,5} Y_{d-1}^{\min} + \beta_{h,6} Y_{d-1}^{\max}}_{\text{non-linear effects}} + \underbrace{\beta_{h,7} Y_{d-1,24}}_{\text{end-of-day effect}} \\
& + \underbrace{\sum_{j=1}^{7} \beta_{h,j+7} D_j}_{\text{weekday dummies}} + \varepsilon_{d,h}
\end{aligned}
$$

- where:
  - $Y_{d,h}$ refers to observed spot price on day d, hour h
  - D_j is a dummy variable for day $j = 1, \ldots, 7$

Benchmark #2: NAR neural net model



- The same inputs as for benchmark #1
- Nonlinear (hidden layer)
- Averaged over 5 runs (→ see the talk by Grzegorz Marcjasz)

Error measures

- Mean Absolute Error:

$$\mathrm{MAE} = \frac{1}{24D} \sum\nolimits_{d=1}^{D} \sum\nolimits_{h=1}^{24} \left|\widehat{\epsilon}_{d,h}\right|$$

- Root Mean Squared Error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{24D} \sum\nolimits_{d=1}^{D} \sum\nolimits_{h=1}^{24} \widehat{\epsilon}^{2}_{d,h}}$$

- To aggregate results across 12 datasets we define the mean percentage deviation from best:

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

Results

MAE and m.p.d.f.b. for the 'expert' model

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	m.p.d.f.b. 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%

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

MAE and m.p.d.f.b. for the neural net

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	m.p.d.f.b. RMSE
original	6.385	4.027	5.447	5.044	4.374	5.716	5.191	1.830	6.042	6.198	4.698	8.144	3.50%	3.42%
$3\sigma_1$	6.203	**3.982**	**5.305**	4.970	4.334	5.314	4.984	1.817	6.383	6.559	4.689	9.320	3.97%	5.31%
$3\sigma_2$	6.252	3.989	5.329	5.029	4.339	5.444	5.036	1.794	6.065	6.231	4.708	8.287	2.36%	2.80%
$3\sigma\log_1$	6.250	4.012	5.347	4.992	4.350	5.353	5.036	1.803	6.073	6.211	4.710	7.974	1.92%	2.09%
$3\sigma\log_2$	6.288	4.011	5.372	5.013	4.357	5.474	5.101	1.818	6.061	6.231	4.709	8.047	2.52%	2.31%
logistic_1	6.205	4.046	5.382	4.998	4.372	5.296	4.972	1.830	6.346	6.502	4.735	8.346	3.21%	5.13%
logistic_2	6.205	4.022	5.338	4.999	4.346	5.463	5.001	1.810	6.287	6.441	4.715	8.322	3.04%	4.30%
asinh_1	6.192	4.011	5.331	4.974	**4.310**	**5.253**	4.936	**1.775**	6.092	6.209	4.672	7.563	0.74%	1.08%
asinh_2	**6.189**	4.012	5.343	**4.956**	4.313	5.384	4.989	1.781	6.086	6.214	**4.659**	7.808	1.31%	1.45%
boxcox_1	6.256	4.008	5.382	5.006	4.320	5.360	5.042	1.785	6.007	6.184	4.697	7.628	1.34%	1.37%
boxcox_2	6.268	4.021	5.400	5.007	4.335	5.526	5.077	1.802	6.033	6.211	4.691	7.783	2.08%	1.98%
poly_1	6.261	4.002	5.355	4.996	4.314	5.341	5.007	1.785	6.019	6.165	4.690	7.665	1.19%	1.37%
poly_2	6.259	4.018	5.378	5.000	4.331	5.485	5.080	1.797	6.035	6.192	4.691	7.799	1.91%	1.68%
mlog_1	6.227	4.013	5.368	4.990	4.332	5.322	4.998	1.784	6.037	6.158	4.693	7.701	1.22%	1.41%
mlog_2	6.240	4.007	5.384	5.009	4.324	5.475	5.062	1.796	6.055	6.171	4.681	7.838	1.86%	1.85%
N-PIT	6.219	4.003	5.308	4.994	4.332	5.264	**4.910**	1.789	5.989	6.110	4.680	7.401	**0.39%**	**0.36%**
t-PIT	6.292	4.071	5.355	5.042	4.395	5.317	4.958	1.834	**5.952**	**6.106**	4.721	**7.384**	1.28%	1.22%

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

Testing for significance: Diebold–Mariano

- Define the error function as ($q = 1$ or 2):

$$L^q(\varepsilon_d) = ||\varepsilon_d||_q = \sum_{x=1}^{11} \sum_{h=1}^{24} |P_{d,h} - \hat{P}_{d,h}^{transf_x}|^q$$

  across 11 (European) datasets and 24 hours

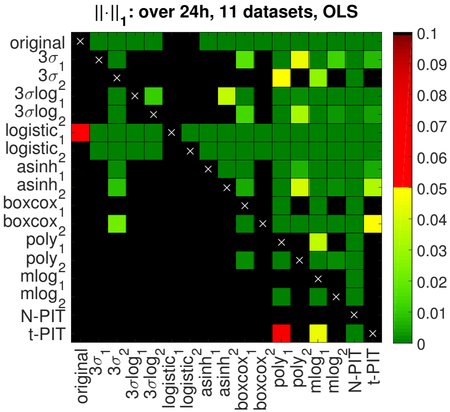
- For each pair of transformations compute the loss differential

$$D_d = L(\varepsilon_d^{transf_X}) - L(\varepsilon_d^{transf_Y})$$

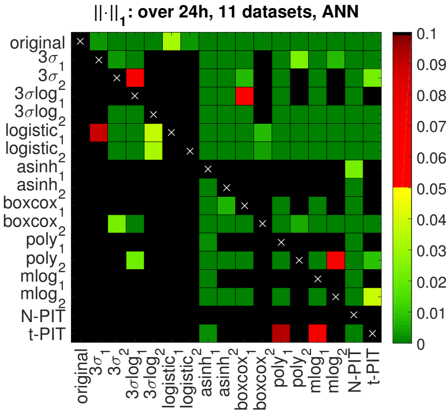
- $H_0 : E(D_d) \leqslant 0$, forecasts for $transf_X$ outperform those for $transf_Y$
- $H_0^R : E(D_d) \geqslant 0$, i.e., the reverse hypothesis

Diebold–Mariano test: *p*-values

'Expert' (AR) model:



Neural net (ANN):



The closer are the *p*-values to zero (dark green) the more significant is the difference between the forecasts of a model on the X-axis (→ better) and those of a model on the Y-axis (→ worse)

Conclusions

- **N-PIT** yields very robust results and significantly outperforms other transformations
  - **mlog** and **asinh** ($\rightarrow$ NAR) are also not bad
- But it is impossible to choose one 'optimal' transformation for all datasets
- In terms of 'normalization' parameters there is no clear indication whether to choose set_1 over set_2 or vice versa ...
  - ... but we recommend to use the more robust set_1, i.e., the (median, MAD) pair

This study can be expanded in several directions ...

- Only two, relatively parsimonious benchmark models considered
- Uniejewski et al. (2017/18, TPWRS) conjecture that the same holds for more complex, e.g., LASSO-estimated models
- Let's see if it really does ...

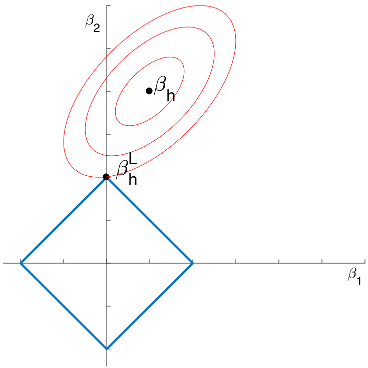

WORK
IN PROGRESS

Least Absolute Shrinkage and Selection Operator

(Tibshirani, 1996, JRSSB)

Minimize the residual sum of squares + a linear penalty function of the β_j's:

$$\hat{\boldsymbol{\beta}} = \underset{\beta_j}{\operatorname{argmin}} \left\{ \underbrace{\sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} \beta_j x_{i,j} \right)^2}_{RSS} + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{penalty} \right\}$$



Blue square – constraint region $|\beta_1| + |\beta_2| \leqslant t$

Red ellipses – contours of the LS error function

Study setup

- Only **one** dataset – Nord Pool
  - System electricity price
  - System consumption *prognosis*
  - Wind power *prognosis* for Denmark
  - All three deseazonalized using wavelet smoothing, as suggested by Marcjasz, Uniejewski & Weron (2018, IJF)
- **Two** baseline autoregressive models: $Lasso_1$ and $Lasso_2$, respectively with **104**(4) and **394**(192) explanatory variables*
- **4** out of 16 best performing VSTs

*(x) indicates the number of eXogenous variables used

Preliminary results

- Benchmark: 'expert' ARX model with one eXogenous variable
- L^1-type 'normalization' set_1: (Median, MAD)
- A more complex model $\longrightarrow$ larger differences
- **N-PIT** and especially **asinh** perform well ...
- ... like for the neural net

MAE

	ARX	Lasso_1	Lasso_2
original	1.7419	1.60032	1.55664
asinh	1.649	1.44833	1.34138
poly	1.6398	1.4546	1.38432
mlog	1.6387	1.44589	1.37287
NPIT	1.6455	1.46597	1.34396

MSE

	ARX	Lasso_1	Lasso_2
original	16.159	14.0336	14.6368
asinh	15.52	13.5382	11.43
poly	14.721	12.0162	12.2902
mlog	14.757	12.0054	12.8548
NPIT	15.485	14.0755	11.5386

energies

Energies **2016**, *9*, 621; doi:10.3390/en9080621 MDPI

Article OPEN ACCESS

Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting

Bartosz Uniejewski, Jakub Nowotarski and Rafał Weron *

Department of Operations Research, Wrocław University of Technology, 50-370 Wrocław, Poland;
uniejewskibartosz@gmail.con
* Correspondence: rafal.we

Academic Editor: Javier Cont
Received: 5 July 2016; Accepte

Abstract: In day-ahead e
Conducting an empirical st
datasets from three major
procedures (single-step eli
we show that using the
commonly-used EPF mode
in EPF before, stands out a

IEEE Xplore® Digital Library

IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. XX, NO. X, 2017/2018

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 (VSTs) 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*** **(PIT) combined with the standard Gaussian distribution yields the best approach, significantly better than many of the considered alternatives.**

$\hat{F}_Z$	Estimate or distributional forecast of F_Z
$\hat{\varepsilon}_{X,d}$	Vector of 24 hourly out-of-sample errors for day d of VST X, i.e., $(\hat{\varepsilon}_{X,d,1},\ldots,\hat{\varepsilon}_{X,d,24})'$
$\|\cdot\|_p$	p-norm, i.e., $\|\hat{\varepsilon}_{X,d}\|_p = (\sum_{h=1}^{24} \|\hat{\varepsilon}_{X,d,h}\|^p)^{1/p}$
$\Delta_{X,Y,d}$	Loss differential series of the multivariate Diebold-Mariano (DM) test, see Eqn. (18)
$\Delta_{X,Y,d,M}$	Loss differential series of the multivariate DM test across 11 European markets (M_1,...,M_{11}), see Eqn. (19)

I. Introduction

Over the last two decades *short-term* (also called *spot* or *day-ahead*; for a discussion see [1]) *electricity price forecasting* (EPF) has joined short-term load forecasting as one

Bibliography

1. B. Uniejewski, J. Nowotarski, R. Weron (2016) *Automated variable selection and shrinkage for day-ahead electricity price forecasting*, Energies 9(8), 621 (doi: 10.3390/en9080621)
2. B. Uniejewski, R. Weron, F. Ziel (2018) *Variance stabilizing transformations for electricity spot price forecasting*, IEEE Trans. Power Systems (doi: 10.1109/TPWRS.2017.2734563)

3. G. Diaz, E. Planas (2016) *A note on the normalization of Spanish electricity spot prices*, IEEE Transactions on Power Systems 31(3), 2499–2500
4. S. Schneider (2011) *Power spot price models with negative prices*, Journal of Energy Markets 4(4), 77-102
5. R. Weron (2014) *Electricity price forecasting: A review of the state-of-the-art with a look into the future*, International Journal of Forecasting 30(4), 1030-1081
6. F. Ziel (2016) *Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure*, IEEE Trans. Power Systems 31(6), 4977–4987
7. F. Ziel, R. Weron (2018) *Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks*, Energy Economics, *forthcoming*. Earlier version available from RePEc: https://ideas.repec.org/p/wuu/wpaper/hsc1608.html