Electricity Price Forecasting with Principal Component-Guided Sparse Regression

Takuji Matsumoto
Faculty of Transdisciplinary Sciences for Innovation
Kanazawa University
Kanazawa, Japan
mtakuji@staff.kanazawa-u.ac.jp

Florian Ziel
House of Energy Markets and Finance
University of Duisburg–Essen
Essen, Germany
Florian.Ziel@uni-due.de

Abstract— **This study investigates forecasting day-ahead electricity prices in markets influenced by weather-dependent energy sources. We utilize the principal component-guided sparse regression (pcLasso) technique, developed by Tay et al., to manage multicollinearity among weather and grid variables effectively. Our focus is on the Japanese Electricity Power Exchange (JEPX), which integrates significant solar capacity, demonstrating how various weather variables impact electricity prices. Our results show that pcLasso surpasses state-of-the-art models like Lasso and Ridge regression in stability and consistency of parameter estimation. Importantly, pcLasso enhances forecast accuracy as more variables are incorporated, providing insights into the seasonality and inter-annual variations of weather impacts. These findings underscore pcLasso's ability to generate robust, interpretable models for complex forecasting scenarios. A key finding when applying pcLasso to JEPX price is the significant improvement in accuracy using day-ahead reserve margin forecasts over weekly forecasts, highlighting the importance of timely disclosure of grid information.**

Index Terms-- **electricity price forecasting, pcLasso, renewable energy sources, weather forecasting**

I. INTRODUCTION

In competitive electricity markets, day-ahead price forecasting is essential for electricity trading practices such as efficient generation and procurement planning and risk management by market participants. In recent years, the penetration of intermittent renewable energy sources such as solar power has increased supply and demand uncertainty, which has had a significant impact on the formation of wholesale electricity prices [1]. Given that much of the variation in supply and demand depends on weather conditions, clarifying how multiple weather conditions simultaneously affect electricity prices would be an interesting research question to understand the electricity market in recent years, where weather-related risks have increased. In addition, since weather effects should vary seasonally and over time, visualizing their occurrence would have practical implications for electricity trading practices. Furthermore, exploring the methodological refinements could contribute to the technical aspects of model development. In this sense, the objective of this study is to explore a highly interpretable and accurate electricity price forecasting model using multiple weather forecast and grid information, and to clarify its effectiveness through extensive empirical analysis.

Many models have been proposed in the context of electricity price forecasting [2][3], often using forecast information about the power system as exogenous variables. For example, some use electricity demand and reserve margins [4][5], and others use forecast information on renewable energy generation such as solar and wind [6][7]. Another approach is to construct a price model using weather information instead of grid information, based on the idea that demand and renewable energy production are strongly related to weather conditions, and there are some previous studies, albeit limited, on this approach as well[1]. However, estimating a robust price model is not an easy task due to the strong multicollinearity among weather variables, and several previous studies [10]–[12] have only used temperature as a weather variable.

When it comes to studies that deal with multiple weather forecasts simultaneously and try to identify their impact, the examples are more limited and, to our knowledge, only [13] and [14]. Huurman et al [13] estimated models combined with ARIMA and GARCH using temperature and wind speed forecasts for electricity prices in Norway and Denmark and showed the predictive power of this weather information in terms of forecast error reduction. Sgarlato and Ziel [14] analyzed the impact of multiple weather forecasts, such as wind speed and solar radiation, on German market prices. Using the variable selection approach of Lasso regression [15], they present a suggestive analysis of the degree of influence of

[1] In relatively recently liberalized markets such as Japan, there is insufficient grid forecast information, so weather forecasts are often used for price forecasting, such as in [8] and [9], but these studies do not focus on weather information.

This research is funded by Grants-in-Aid from the Japan Society for the Promotion of Science (JSPS) – 21K14374, 22KK0217, and 24K16396 (to T.M.), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 505565850 (to F.Z.).

979-8-3503-8174-0/24/$31.00 ©2024 IEEE

each weather information. Although their interest was in identifying the role of weather forecasts in the horizon beyond the day-ahead, and thus focused on a specific period, their weather variable selection approach would be helpful for the purpose of this study, i.e., to capture (medium- and long-term) seasonal and interannual changes in weather impacts.

Therefore, in this study, based on the analytical approach of [14], we consider applying a more robust methodology to allow for stable model estimation across all seasons. The Lasso regression used in their research, along with Ridge regression [16] and others, is classified as sparse regression. Sparse regression is a model suitable for building highly interpretable models as it can eliminate (or shrink) redundant variables and extract essential variables. Regarding Ridge regression, a meta-analysis [17] showed that the stronger the multicollinearity in the explanatory variables, the higher the predictive accuracy of Ridge regression than general OLS, etc. As for other sparse regressions, several extended models with various innovations have been proposed in recent years [18]–[21]. Among them, principal component-guided sparse regression (pcLasso), developed by Tay et al. [18], is reported to be excellent for extracting variables from feature-rich data and more accurate than similar conventional methods in the literature. As pcLasso is a relatively recently proposed method, there are few examples of its application to empirical studies. To the best of our knowledge, it has only been applied to data analysis in bioinformatics [22][23] and cryptocurrency price analysis [24][25]. Thus, we believe that applying pcLasso to energy market analysis and forecasting models will be a new research contribution in terms of expanding the applications of the method.

For the empirical analysis, we choose the Japanese market, where solar power generation is increasingly installed. In markets with high solar generation, weather factors that affect electricity supply (e.g., temperature and rainfall) also have a significant impact on the demand side (albeit differently in different seasons), so the way weather affects price would be more complex than in markets with high wind generation. Since a robust estimation of the price model is assumed to be difficult, an empirical analysis using data from the Japanese market is considered an ideal case study to validate the effectiveness of the model. With this in mind, this study applies pcLasso to JEPX spot price forecasts to visualize the degree of influence of various weather and other information, and to perform a multi-faceted verification of the model's validity and predictive accuracy. This study is organized as follows. The next section provides an overview of the pcLasso model, followed by an empirical analysis in Section III. Finally, Section IV presents the conclusions.

II. METHOD

A. pcLasso Model

In the pcLasso model, when the explanatory variable matrix $\boldsymbol{X}$ is singular value decomposed as $\mathbf{X} = \mathbf{UDV}^T$ (where $\mathbf{D}$ is a diagonal matrix whose elements are singular values $d_1 \geq d_2 \dots \geq d_m > 0, m = \mathrm{rank}(\mathbf{X})$), consider the problem of minimizing $J(\boldsymbol{\beta})$ in the following formula [18].

$$J(\boldsymbol{\beta}) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 + \frac{\theta}{2}\boldsymbol{\beta}^T\mathbf{V}\mathbf{D}_{d_1^2-d_j^2}\mathbf{V}^T\boldsymbol{\beta} \qquad (1)$$

where $\mathbf{y}$ is the vector of objective variables, $\boldsymbol{\beta}$ is the parameter (coefficient) vector to be estimated, $\mathbf{D}_{d_1^2-d_j^2}$ is an $m \times m$ diagonal matrix with $d_1^2 - d_j^2 (j = 1, \dots, m)$ as elements, and λ and θ are parameters that adjust the strength of the penalty (regularization) in the second and third terms. Up to the second term of (1), it is equivalent to Lasso regression, and the second term is the usual L1 penalty. On the other hand, the characteristic third term is derived from the $\boldsymbol{\beta}^T\mathbf{VZV}^T\boldsymbol{\beta}$ form (where $\boldsymbol{Z}$ is a diagonal matrix whose diagonal elements are functions of the squared singular values), a generalization of the L2 (quadratic) penalty, and is a more elaborate variant of it. The idea is to penalize each component (i.e., the original coefficients that synthesize each component) according to the deviation of its eigenvalue from the principal components. In other words, by weighting the penalties so that the coefficients of variables with smaller contributions to the principal component are more likely to be shrunk, the extraction of essential variables is made more efficient.

B. Parameter Estimation Model

In this study, we set the objective variable as the daily average electricity day-ahead price S_t (on day t) and estimate the following model:

$$\begin{aligned} S_t = {} & \beta_0 + \beta_1 Sin_{1,t} + \beta_2 Cos_{1,t} + \beta_3 Sin_{2,t} + \beta_4 Cos_{2,t} \\ & + \beta_5 t + \beta_6 Holiday_t + \beta_7 Saturday_t \\ & + \beta_8 Monday_t + \beta_9 Tmr_t + \beta_{10} {Tmr_t}^2 \\ & + \beta_{11} Trange_t + \beta_{12} Rain_t + \beta_{13} Margin_t \\ & + \textstyle\sum_{i=1}^{14} \beta_{14+i} S_{t-i} + \varepsilon_t \end{aligned} \qquad (2)$$

where $Sin_{k,t} := \sin\left(\frac{2\pi kt}{365.25}\right), Cos_{k,t} := \cos\left(\frac{2\pi kt}{365.25}\right)$; $Holiday_t$ (which includes national holidays and Sundays), $Saturday_t$, and $Monday_t$ are day-type dummy variables; Tmr_t and $Trange_t$ are mid-range and range of maximum-minimum temperature forecasts, i.e., $Tmr_t \coloneqq (Tmax_t + Tmin_t)/2$, $Trange_t \coloneqq Tmax_t - Tmin_t$; $Rain_t$, $Margin_t$ are the rainfall probability and reserve margin forecast, respectively; β is coefficient, and ε_t is the residual. Note that for the mid-range temperature, the quadratic term ${Tmr_t}^2$ is included based on the results in [26].

In this study, (2) is estimated using OLS, Lasso, Ridge, and pcLasso, and the coefficients and prediction errors are compared. We standardize all data sets used in each estimation (returning to the original series when checking the prediction error). This not only gives equal weight to the penalties of the features, but also makes it possible to compare the influence of (different units of) explanatory variables when validating the estimated parameters. Note that Lasso and Ridge estimate the regularization parameter λ with 5-fold cross-validation. The pcLasso also estimates λ with 5-fold cross-validation, but we additionally assume six different "$ratio$" parameters (related to θ; see [18]): $ratio \in \{0.25, 0.5, 0.75, 0.9, 0.95, 0.99\}$, and choose the combination of λ and $ratio$ that yields the smallest mean squared error.

III. RESULT

The data used in this study for the empirical analysis are as follows and each explanatory variable after processing are plotted them in Fig. 1.

- Electricity day-ahead (spot) price [JPY/kWh]: JEPX spot price (Tokyo area) [2]
- Maximum and minimum temperature [°C], rainfall probability: Day ahead weather forecast by JMA[3]
- Reserve margin [%]: Weekly reserve margin published by OCCTO [4]

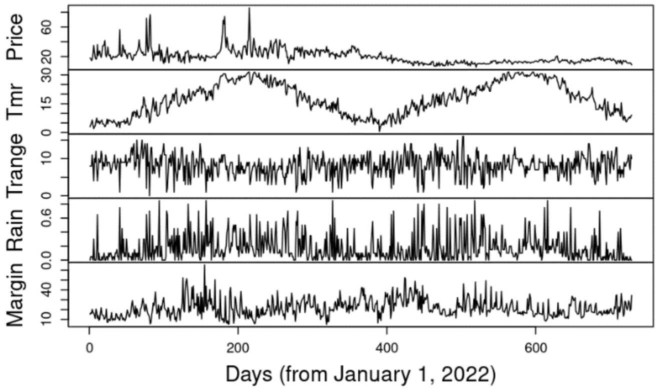


Fig. 1: Data description

For the out-of-sample period, we use a two-year period from January 1, 2022 to December 31, 2023 to include periods of extreme electricity price spikes and calm periods in the validation process, and for the forecast for each day t, we use observed data for the past 42 days (i.e., daily rolling forecasts are made using data for the in-sample period $[t-42, t-1]$). Note that in the Japanese market, in addition to the weekly reserve margin, the day-ahead reserve margin is also published. However, the day-ahead reserve margin is published in the evening of the previous day and cannot be used for bidding in the spot market, which is conducted in the morning, so the weekly reserve margin is used in this study. In addition, as a hypothetical case, we also measure the forecast error if the day-ahead reserve margin could be used for forecasting, and the results are presented at the end of Section III-B.

A. Model Estimation Results

First, to confirm the rationality of the estimation results of the pcLasso model, we plot the coefficients of temperature-related variables, rainfall probability and weekly reserve margin in Fig. 2. As can be seen, the sign of the temperature coefficient is positive in summer and negative in winter, which is consistent with the increasing/decreasing demand for cooling/heating. The rainfall probability coefficient is generally positive, presumably due to the decrease in solar radiation or solar power generation, but is relatively small in the summer season. This may be due to the additional cooling effect (lower demand) of rainfall. The coefficient of the temperature squared variable tends to be positive especially when the absolute value of the temperature coefficient is large, which is consistent with the fact that the nonlinearity increases when the sensitivity to temperature is high. The coefficients for the reserve margin are generally negative, consistent with the tendency that the smaller the reserve margin, the more likely a price spike will occur. The temperature difference coefficient (Trange) tends to be negative, but tends to approach zero when the reserve margin coefficient is large. This is probably because the temperature difference is related to the amount of solar power generated (solar radiation around noon) and therefore has a smaller effect during periods of higher margin sensitivity (when supply and demand are tight, causing price spikes in the evening).

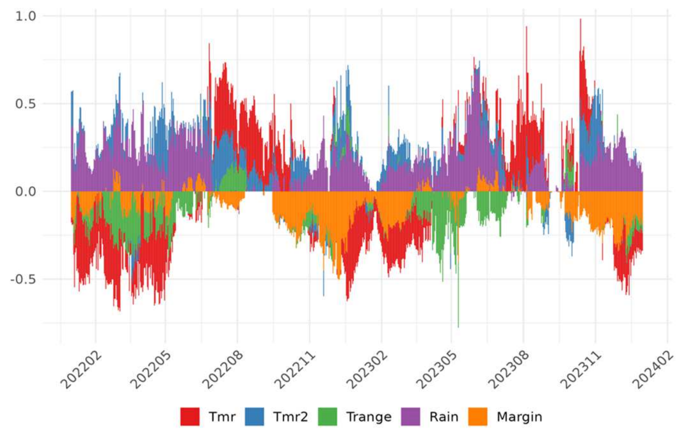


Fig. 2: Changes in estimated parameters

Next, to confirm the validity of pcLasso's estimation results, a comparison of each method is shown in Fig. 3. pcLasso shows the smoothest results with the least fluctuation, followed by Ridge and Lasso, which become increasingly unstable. Although not shown in the graph, the OLS estimation results were incomparably more unstable. This can be interpreted as the unstable parameter estimation in OLS due to multicollinearity among variables is stabilized by the regularization of sparse regressions. And it may also mean that the more the quadratic penalty is refined, the smoother the estimation becomes.

We then extract variables with signs of the estimated parameters that are considered consistent across seasons and summarize in TABLE I the percentage of estimated results that are consistent with these signs (where the percentage is calculated as the area of coefficients with consistent signs relative to the total area of estimated coefficients in Figure 3). From this table, we can also see that the estimation result of pcLasso is highly consistent.

[2] Downloaded from http://www.jepx.org/market/index.html.

[3] Downloaded from https://pe-sawaki.com/WeatherForecast/. Note that JMA is Japan Mereological Agency, Japan.

[4] Downloaded from https://occtonet3.occto.or.jp/public/dfw/RP11/OCCTO/SD/LOGIN_login#. We use a 7-day forecast starting the following Saturday, published every Thursday. Note that OCCTO is Organization for Cross-regional Coordination of Transmission Operators, Japan.

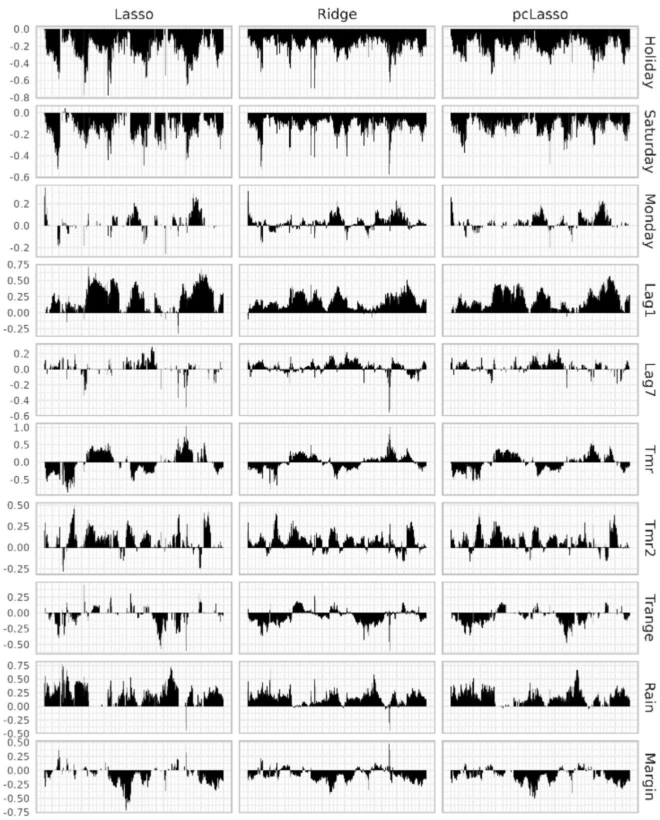


Fig. 3: Changes in estimated parameters by methods

TABLE I. COMPARISON OF PARAMETER SIGN CONSISTENCY BETWEEN METHODS

Variables	Plausible sign	OLS	Lasso	Ridge	pcLasso
Holiday	Negative	99.6%	**100.0%**	**100.0%**	**100.0%**
Saturday	Negative	96.7%	99.9%	**100.0%**	**100.0%**
Monday	Positive	36.4%	83.3%	81.4%	**88.7%**
Lag1	Positive	30.0%	99.3%	99.8%	**99.9%**
Lag7	Positive	7.8%	68.2%	75.2%	**84.2%**
Rain	Positive	96.7%	99.5%	98.8%	**99.8%**
Margin	Negative	45.0%	91.0%	87.9%	**92.8%**

B. *Validation of Forecast Error*

Next, we examine the predictive accuracy: Fig. 4 compares the changes in the out-of-sample root mean squared error (RMSE) and mean absolute error (MAE) (divided by the average actual price) as the number of explanatory variables is gradually increased, comparing five models from M1 to M5 across the three sparse regression models (see TABLE II for the correspondence between models and explanatory variables). The results show that pcLasso tends to reduce the prediction error with each increase in the number of explanatory variables. Also, the prediction accuracy of pcLasso tends to be higher per month, as shown in TABLE III, followed by Ridge, Lasso, and OLS.

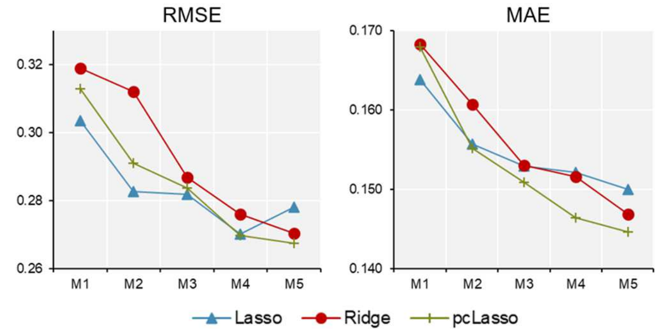


Fig. 4: Comparison of model performance using RMSE and MAE

TABLE II. CORRESPONDENCE WITH EXPLANATORY VARIABLES FOR EACH MODEL (M1-M5)

Variables		M1	M2	M3	M4	M5
–	($Sin_{k,t}, Cos_{k,t}, t, Holiday_t, Saturday_t, Monday_t, S_{t-i}$)	✓	✓	✓	✓	✓
Tmr	($Tmr_t + Tmr_t^2$)		✓	✓	✓	✓
Trange	($Trange_t$)			✓	✓	✓
Rain	($Rain_t$)				✓	✓
Margin	($Margin_t$)					✓

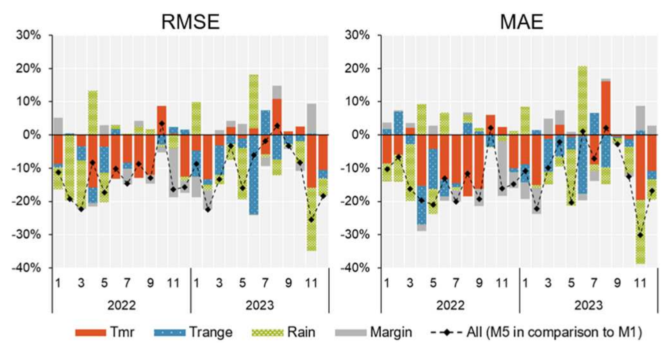
TABLE III. MONTHLY COMPARISON OF MODEL PERFORMANCE USING RMSE AND MAE (M5)

Month	RMSE				MAE			
	OLS	Lasso	Ridge	pcLasso	OLS	Lasso	Ridge	pcLasso
202201	1.769	**0.267**	0.290	0.276	0.532	**0.216**	0.235	0.223
202202	0.462	0.249	0.256	**0.246**	0.340	0.168	0.170	**0.161**
202203	0.597	0.505	**0.416**	0.426	0.461	0.338	**0.276**	0.279
202204	0.734	0.434	0.381	**0.380**	0.520	0.336	0.285	**0.282**
202205	0.473	0.191	0.185	**0.178**	0.402	0.144	**0.135**	0.136
202206	0.265	0.223	0.226	**0.215**	0.155	0.126	**0.124**	0.126
202207	0.980	**0.291**	0.336	0.319	0.670	**0.167**	0.195	0.180
202208	0.809	**0.288**	0.311	0.331	0.599	**0.164**	0.168	0.185
202209	0.395	0.200	0.231	**0.199**	0.279	0.160	0.181	**0.150**
202210	0.493	0.208	0.213	**0.201**	0.372	**0.151**	0.157	0.160
202211	0.186	0.113	0.113	**0.110**	0.146	0.089	0.092	**0.088**
202212	0.162	0.108	**0.094**	0.102	0.133	0.085	**0.077**	0.084
202301	0.657	0.112	0.107	**0.106**	0.253	0.096	0.088	**0.088**
202302	0.222	0.117	**0.103**	0.108	0.165	0.086	**0.075**	0.078
202303	0.311	0.159	0.148	**0.144**	0.248	0.125	0.118	**0.116**
202304	0.330	**0.215**	0.223	0.227	0.293	**0.176**	0.185	0.189
202305	0.278	0.187	0.174	**0.171**	0.217	0.150	0.133	**0.132**
202306	0.226	0.159	**0.135**	0.146	0.162	0.116	**0.100**	0.108
202307	0.183	**0.091**	0.100	0.100	0.152	**0.069**	0.076	0.077
202308	0.126	**0.064**	0.070	0.067	0.106	0.051	0.053	**0.050**
202309	0.226	**0.100**	0.105	0.105	0.167	**0.077**	0.081	0.081
202310	0.359	**0.118**	0.143	0.123	0.251	**0.095**	0.113	0.100
202311	0.184	0.056	**0.054**	0.059	0.148	0.046	**0.044**	0.047
202312	0.185	0.092	0.091	**0.091**	0.162	0.077	0.077	**0.076**
All period	0.718	0.278	0.270	**0.267**	0.324	0.150	0.147	**0.145**

Looking at the rates of reduction in RMSE and MAE for the M5 model relative to the M1 model, Fig. 5 shows that the combination of weather and other variables that contribute to the improvement in forecast error varies from month to month. In 2022, when the reserve margin was relatively small (due to high fossil fuel prices and supply shortages) and price fluctuations were relatively large, the reduction in forecast error due to temperature was large, but in 2023, when supply and demand were slack and prices were stable, temperature differences (related to solar power generation) contributed more significantly to the reduction in error.

Fig. 6 plots the forecast error reduction rate of M5 relative to M4 (i.e., the error reduction rate due to the weekly reserve margin) as a gray bar (corresponding to the gray part of the bar in Fig. 5) and the error reduction rate due to the reserve margin in the model where the weekly margin is replaced by the day-ahead margin as a yellow bar (as mentioned in Section III, this is a hypothetical case because the day-ahead margin is published after the spot market bids and therefore cannot be used for actual forecasting). Of the 24-month periods, the weekly margin contributed to reducing the RMSE (MAE) in 14 (12) months, while the day-ahead margin reduced the forecast error in more cases (18 (15) months) and in some months the RMSE and MAE improved by almost 20%.

Since reserve margin information reflects specific generator operating plans, etc., more immediately prior to bidding, updating such information can improve the accuracy of the price forecast. In the Japanese market, as mentioned above, the only reserve margin information that can be referenced for day-ahead bidding is the weekly forecast, but publication of the latest reserve margin forecast immediately prior to bidding should be considered. Increased transparency of grid information will lead to more accurate price signals for market participants, which in turn will encourage more sophisticated planning of electricity consumption and procurement (and thus contribute to more efficient use of resources). Empirical results such as those presented here, in which reserve margin information updates contribute to improved price forecasting accuracy, may have such policy implications.



Note: Values by explanatory variable are calculated from the reduction rate of RMSE or MAE for M2 to M5 relative to M1 in the pcLasso model.

Fig. 5: RMSE and MAE reduction rate for each exogenous variable

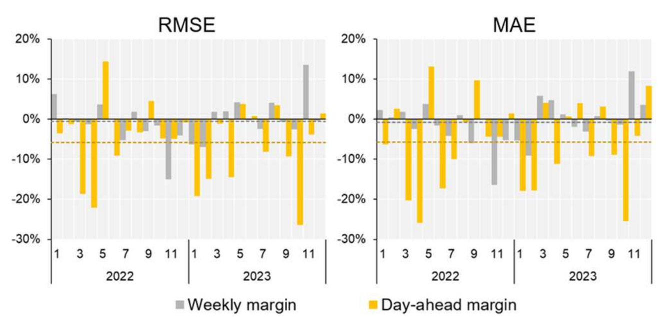


Fig. 6: RMSE and MAE reduction rate by reserve margin forecast (week ahead vs. day ahead)

IV. Conclusion

This study applies pcLasso, recently developed by Tay et al. [18], to electricity price forecasting using weather forecasts, etc., and demonstrates the effectiveness of the model in terms of both rationality and forecast accuracy. In particular, the contributions of this paper are as follows.

- It is shown that pcLasso provides more stable parameter estimation results and better consistency with general interpretations than OLS, Lasso, Ridge, etc., and that it improves out-of-sample forecast errors with each variable included.
- Practical implications for electricity trading are provided, such as visualizing the seasonality and interannual changes in maximum and minimum temperature-related variables, rainfall probability, and reserve margin coefficients, and verifying their appropriateness with different interpretations.
- In addition to finding that weekly reserve margin information improves price forecasting error, the study also found that forecast accuracy could have been further improved if the day-ahead reserve margin had been disclosed, adding to the policy discussion on the need to improve information transparency.

pcLasso has the potential to estimate relatively robust and stable models (reasonable parameters for each variable) even when there is collinearity among explanatory variables (e.g., lag variables, temperatures, rainfall probability, etc.) or when explanatory variables that are not strongly related are included (e.g., squared temperature terms or weekly reserve margin). In addition, sparse regression including pcLasso is easier to interpret and more computationally efficient than many machine learning methods, making it an easy model to handle in practice (e.g. decision making in electricity trading). The ability of pcLasso to effectively incorporate multiple explanatory variables and still produce robust estimates is a major advance in modeling technology, and sophisticated models based on it can be expected to contribute to improved decision making in electricity markets.

Acknowledgment

This research is funded by Grants-in-Aid from the Japan Society for the Promotion of Science (JSPS) – 21K14374, 22KK0217, and 24K16396 (to T.M.), and the Deutsche Forschungs-gemeinschaft (DFG, German Research Foundation) – 505565850 (to F.Z.).

References

[1] A. Gianfreda and D. W. Bunn, “A stochastic latent moment model for electricity price formation,” Operations Research, vol. 66, no. 5, pp. 1189–1203, 2018.

[2] 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.

[3] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, “Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark,” Applied Energy, vol. 293, 116983, 2021.

[4] L.I. Hagfors, D. W. Bunn, E. Kristoffersen, T. T. Staver, and A. Westgaard, S, “Modeling the UK electricity price distributions using quantile regression,” Energy, vol. 102, pp. 231–243, 2016.

[5] N. V. Karakatsani and D. W. Bunn, “Forecasting electricity prices: The impact of fundamentals and time-varying coefficients,” International Journal of Forecasting, vol. 24, no. 4, pp. 764–785, 2008.

[6] F. Ziel, R. Steinert, and S. Husmann, “Efficient modeling and forecasting of electricity spot prices,” Energy Economics, vol. 47, pp. 98–111, 2015.

[7] G. Marcjasz, B. Uniejewski, and R. Weron, “Beating the naïve—Combining LASSO with naïve intraday electricity price forecasts,” Energies, vol. 13, no. 7, 1667, 2020.

[8] M. Omura, Y. Fujimoto, H. Ishii, Y. Hayashi, T. Sawa, H. Sasaki, and N. Fukuyama, “Day-Ahead Electricity Price Forecasting Using an Adaptive Combination Method in the Japanese Spot Market,” In 2020 17th International Conference on the European Energy Market, pp. 1–6, 2020.

[9] T. Matsumoto, D. W. Bunn, and Y. Yamada, “Pricing electricity day-ahead cap futures with multifactor skew-t densities,” Quantitative Finance, vol. 22, no. 5, pp. 835–860, 2022.

[10] S. Bigerna, “Estimating temperature effects on the Italian electricity market,” Energy Policy, vol. 118, pp. 257-269, 2018.

[11] C. R. Knittel and M. R. Roberts, “An empirical examination of restructured electricity prices,” Energy Economics, vol. 27, no. 5, pp. 791–817, 2005.

[12] T. Matsumoto and M. Endo, “One-week-ahead electricity price forecasting using weather forecasts, and its application to arbitrage in the forward market: An empirical study of the Japan electric power exchange,” Journal of Energy Markets, vol. 14, no. 3, pp. 39–64, 2021.

[13] C. Huurman, F. Ravazzolo, C. Zhou, “The power of weather,” Computational Statistics & Data Analysis, vol. 56, no. 11, pp. 3793–3807, 2012.

[14] R. Sgarlato and F. Ziel “The Role of Weather Predictions in Electricity Price Forecasting Beyond the Day-Ahead Horizon,” IEEE Transactions on Power Systems, vol. 38, no. 3, pp. 2500–2511, 2022.

[15] R. Tibshirani, “Regression shrinkage and selection via the lasso,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

[16] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” Technometrics, vol. 12, no. 1, pp. 55–67, 1970.

[17] H. Duzan and N. S. B. M. Shariff, “Ridge regression for solving the multicollinearity problem: review of methods and models,” Journal of Applied Science, vol. 15, no. 3, pp. 392–404, 2015.

[18] J. K. Tay, J. Friedman, and R. Tibshirani, “Principal component-guided sparse regression,” Canadian Journal of Statistics, vol. 49, no. 4, pp. 1222–1257, 2021.

[19] D. Bertsimas and M. S. Copenhaver, “Characterization of the equivalence of robustification and regularization in linear and matrix regression,” European Journal of Operational Research, vol. 270, no. 3, pp. 931–942, 2018.

[20] P. Jain, N. Rao, and I. S. Dhillon, “Structured sparse regression via greedy hard thresholding,” Advances in neural information processing systems, vol. 29, pp. 1516–1524, 2016.

[21] P. B. Nair, A. Choudhury, and A. J. Keane, “Some greedy learning algorithms for sparse regression and classification with mercer kernels,” Journal of Machine Learning Research, vol. 3, pp. 781–801, 2002.

[22] G. Magazzù, G. Zampieri, and C. Angione, “Multimodal regularized linear models with flux balance analysis for mechanistic integration of omics data,” Bioinformatics, vol. 37, no. 20, pp. 3546–3552, 2021.

[23] A. Kenny, D. Solomon, M. S. Patwary, and K. P. Das, “Sparse regression with clustered predictors,” Journal of Statistical Research, vol. 56, no. 1, pp. 37–53, 2022.

[24] T. Panagiotidis, T. Stengos, and O. Vravosinos, “A principal component-guided sparse regression approach for the determination of bitcoin returns,” Journal of Risk and Financial Management, vol. 13, no. 2, 33, 2020.

[25] P. Deniz and T. Stengos, “Cryptocurrency Returns before and after the Introduction of Bitcoin Futures,” Journal of Risk and Financial Management, vol. 13, no. 6, 116, 2020.

[26] T. Matsumoto, Y. Yamada, “Simultaneous hedging strategy for price and volume risks in electricity businesses using energy and weather derivatives,” Energy Economics, vol. 95, 105101, 2021.