

主成分誘導型スパース回帰を用いた電力スポット価格予測

05000966 金沢大学 *松本 拓史 MATSUMOTO Takuji

1. はじめに

電力スポット市場（前日市場）価格の予測は、効率的な発電・需要計画の策定等の電力取引実務に不可欠である。太陽光発電の普及により、価格に対する気象の連動性が強まっている近年[1]では、価格予測における気象情報の重要性が高まっている[2]。しかし、気象変数間には強い共線性が存在するため、頑健な価格モデルの推定は容易でなく、[1]や[2]の既往研究では、最高気温のみが用いられている。電力系統情報が広く取得できる海外市場では、気象情報を活用した研究は限定的だが、Sgarlato & Ziel[3]が Lasso 回帰による気象予測情報の影響度分析を行っている。

Lasso や Ridge に代表されるスパース回帰は、冗長な変数を除去/縮小し、本質的な変数を抽出できる点で解釈性に優れ、実務での意思決定にも活用しやすい。特に、最近、Tay ら[4]によって開発された主成分誘導型スパース回帰 ("pcLasso") は、特徴量の多いデータの変数抽出に優れているとされ、従来手法よりも精度が高いことが報告されている。そこで、本研究では、JEPX スポット価格予測に pcLasso を適用し、複数の気象情報等の影響度を視覚化するとともに、予測精度の改善に関して多角的な検証を行う。

2. 主成分誘導型スパース回帰 (pcLasso) モデル

pcLasso モデルでは、説明変数行列 \mathbf{X} が $\mathbf{X} = \mathbf{UDV}^T$ と特異値分解されるとき（ただし \mathbf{D} は特異値 $d_1 \geq d_2 \dots \geq d_m > 0$, $m = \text{rank}(\mathbf{X})$ を要素を持つ対角行列）、次式の $J(\beta)$ を最小化する問題を考える[4]。

$$J(\beta) = \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 + \frac{\theta}{2} \beta^T \mathbf{VD}_{a_1^2 - a_j^2} \mathbf{V}^T \beta$$

ここで、 \mathbf{y} は目的変数ベクトル、 β は推定パラメータ（係数）、 $\mathbf{D}_{a_1^2 - a_j^2}$ は $d_1^2 - d_j^2$ ($j = 1, \dots, m$) を要素を持つ $m \times m$ の対角行列であり、 λ と θ は、第 2 項と第 3 項の罰則の強度を調整するパラメータである¹。第 2 項までは Lasso 回帰と等しいが、特徴的な第 3 項が意味するのは、主成分からの固有値（寄与率）の乖離に応じて、各成分（を合成する係数）に罰則を課すと

いう考え方である。つまり、寄与度の小さい係数ほど縮小されやすいように重み付けすることで、本質的な変数の抽出を一層効率化する工夫が施されている。

3. 推定結果と予測精度の検証

JEPX スポット価格²（日平均） S_t に対し、前日価格 S_{t-1} 、7 日前価格 S_{t-7} 、受渡前日朝時点の気象庁天気予報³（最高気温 $Tmax_t$ 、最低気温 $Tmin_t$ 及び日中の降水確率 $Rain_t$ ）と週間予備率⁴ $Margin_t$ を説明変数としたモデルを考える（いずれも東京エリア）。ただし、最高・最低気温については、[1]と同様に 2 乗の項 ($Tmax_t^2$ 及び $Tmin_t^2$) を含める。また、スポット価格は[2]で実証された知見を基に、対数化して用いる。予測期間（out-of-sample 期間）は、2016 年 7 月 1 日から 2022 年 6 月 30 日までの 6 年間とし、各日（第 t 日）の予測には、直近 60 日の観測データを用いる（つまり、in-sample 期間 $[t - 60, t - 1]$ のデータを用いて daily rolling 予測を行う）こととする⁵。

まず、pcLasso モデルの推定結果の合理性を確認するため、図 1 に、最高・最低気温（1 次の項）と降水確率、週間予備率の各係数をプロットする⁶。気温の係数の符号（冬に負、夏場に正）は直感と整合するが、最低気温よりも最高気温の係数の方が総じて大きくなっているのは興味深い。これは、通常、最高気温が観測されるのが午後過ぎあり、電力価格（電力需要）の変動が大きくなる時間帯と一致するためであろう。また、降水確率の係数が総じて正であるのは、太陽光発電量の低下によるものと想定されるが、夏季で相対的に小さくなっている。これは、降雨による冷却効果（需要低下）が影響（相殺）している可能性がある。予備率は、小さくなる程、価格高騰が生じやすいため、係数が総じて負となっていることは直感と整合する。

推定されたパラメータの符号の整合性について、手法間で比較した結果を表 2 に示す。前日や 7 日前のスポット価格の係数は正、予備率の係数は負となるのが整合的な結果であるが、この整合頻度は、総じて pcLasso（次点で Ridge）で高くなっている。

¹ R のパッケージ "pcLasso" では、 λ は交差検証から求められ、 θ は指定する必要がある。本研究では、 θ を決定する変数 "ratio" を 0.95 と設定した。

² <http://www.jepx.org/market/index.html>

³ <https://pe-sawaki.com/WeatherForecast/>

⁴ https://occtonet3.occto.or.jp/public/dfw/RP11/OCCTO/SD/LOGIN_login#

⁵ 推定の都度、すべての説明変数及び目的変数は標準化して用いる。

⁶ ここでは、後述する M5 モデルの推定結果を示しており、該当変数のパラメータのみを表示している。なお、推定結果を推定期間の中点に対応させるため、横軸は t ではなく、 $t - 30$ になるようにプロットしている。

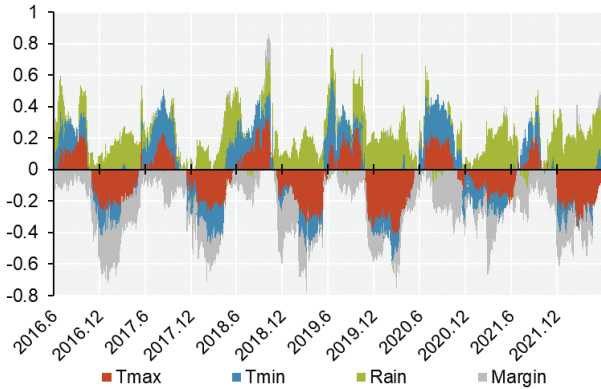


図1：推定パラメータの推移

表1：パラメータ符号の整合性(頻度)の手法間比較

Variables	Plausible Sign	OLS	Lasso	Ridge	pcLasso
S_{t-1}	Positive	98.4%	96.9%	99.5%	99.8%
S_{t-7}	Positive	80.0%	63.3%	89.8%	87.7%
$Margin_t$	Negative	72.9%	91.3%	83.4%	95.0%

次に、予測精度の検証を行う。図2では、M1からM5までの5つのモデルを考え、説明変数を徐々に増やした場合に out-of-sample の RMSE (ただし、平均実績価格で除算) がどう変化するかを手法別に比較している (モデルと説明変数の対応は表2を参照)。この結果が示す通り、pcLasso は説明変数を増やすごとに、RMSE を縮小させている。Ridge も次点で小さいが、最低気温や週間予備率の追加に対しては、予測誤差を拡大させている (なお、MAE でも同様の結果であったが、本稿では紙面の都合上割愛する)。

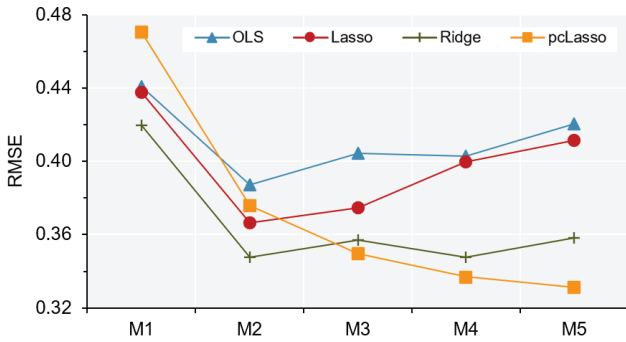


図2：RMSEのモデル別手法間比較

表2：各モデル (M1~M5) の説明変数との対応

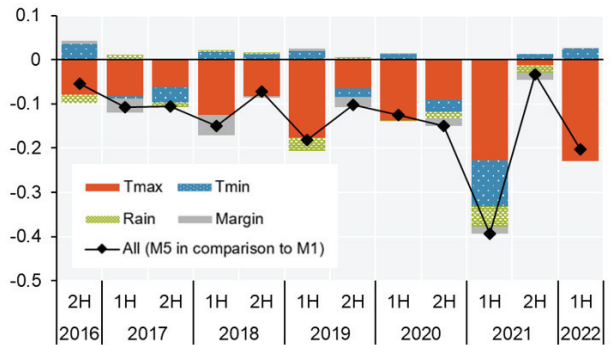
Variables	M1	M2	M3	M4	M5
- ($S_{t-1} + S_{t-7} + Holiday_t$)	✓	✓	✓	✓	✓
Tmax ($Tmax_t + Tmax_t^2$)		✓	✓	✓	✓
Tmin ($Tmin_t + Tmin_t^2$)			✓	✓	✓
Rain ($Rain_t$)				✓	✓
Margin ($Margin_t$)					✓

期間別 (半年毎) に見ても、表3に示す通り、pcLasso の予測精度の高さは安定しており、その次に Ridge, Lasso, OLS と続く結果となっている。

表3：RMSE と MAE の期間別手法間比較 (M5)

Period	RMSE				MAE			
	OLS	Lasso	Ridge	pcLasso	OLS	Lasso	Ridge	pcLasso
2016.H2	0.139	0.140	0.143	0.147	0.100	0.099	0.099	0.102
2017.H1	0.110	0.109	0.109	0.107	0.081	0.080	0.079	0.077
2017.H2	0.191	0.188	0.178	0.179	0.121	0.119	0.115	0.115
2018.H1	0.142	0.141	0.139	0.137	0.088	0.087	0.087	0.086
2018.H2	0.204	0.197	0.199	0.195	0.118	0.114	0.112	0.110
2019.H1	0.225	0.246	0.227	0.213	0.089	0.089	0.089	0.086
2019.H2	0.226	0.222	0.210	0.206	0.132	0.132	0.130	0.129
2020.H1	0.097	0.096	0.095	0.096	0.078	0.076	0.076	0.076
2020.H2	0.691	0.680	0.577	0.589	0.269	0.266	0.243	0.241
2021.H1	0.711	0.689	0.584	0.513	0.253	0.253	0.235	0.217
2021.H2	0.225	0.224	0.222	0.221	0.130	0.126	0.125	0.124
2022.H1	0.354	0.349	0.304	0.288	0.201	0.199	0.190	0.185
All period	0.421	0.412	0.358	0.331	0.149	0.148	0.142	0.138

M1 に対する M5 モデルの RMSE 低減率について見ると、図3に示すとおり、気象情報等の利用による予測誤差の改善幅が、年々大きくなっている傾向が確認される。最高気温の利用のみでも予測誤差は大きく低下するが、価格が大きく高騰した2021年上半期などでは、他の予測情報との組み合わせが、誤差の低減に大きく寄与していることが分かる。



注：説明変数別の値は、M2~M5のM1に対するRMSE低減率から計算している。

図3：M1に対するM5のRMSE低減率 (pcLasso)

4. おわりに

本研究では、気象予測等を用いた電力価格予測に pcLasso を活用し、モデルの合理性と予測精度の両面から有効性を明らかにした。気象情報の地点や種類を増やすなどすれば一層の精度改善余地もあり、電力取引実務での活用と更なる精緻化が期待されよう。

参考文献

- [1] 松本, 遠藤. わが国の電力先物市場におけるリスクプレミアムの実証分析. 電力中央研究所研究報告 Y20004, 1-26, 2021.
- [2] Matsumoto, T., Endo, M. 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, 14(3), 2021.
- [3] Sgarlato, R., Ziel, F. The Role of Weather Predictions in Electricity Price Forecasting Beyond the Day-Ahead Horizon. IEEE Transactions on Power Systems, 2022.
- [4] Tay, J. K., Friedman, J., Tibshirani, R. Principal component-guided sparse regression. Canadian Journal of Statistics, 49(4), 1222-1257, 2021.