Econometric of Electricity Market

**Project: Forecasting France's Day-ahead Electricity Price
Using LASSO-Based Regression Models**

Authors: Ly Le Thi
Supervisors: Prof. Dr. Florian Ziel, M.Sc. Paul Ghelasi

1 Introduction

This study focuses on advancing day-ahead electricity price forecasting in France amid challenges from the dynamic energy landscape and recent geopolitical events. Using three shrinkage methods (Ridge Regression, LASSO, and Elastic Net), the research explores high-dimensional projections for time series forecasting. These three methods are widely used as regularization techniques in predictive modeling, effectively handling large predictor sets, and mitigating multicollinearity and overfitting. Comparative insights are provided by four benchmark models: The Naive model, the Expert model, the Extended Expert model by adding the last hour, and the Extended Expert Model by adding covariates. Results indicate Elastic Net outperforms all other models with the smallest roots mean square error (RMSE), followed by LASSO and Ridge regression. However, in the practical scenarios of out-of-sample forecasting, Ridge regression outperforms all other forecasting models. The limitations of relying solely on historical data are underscored, particularly by the higher RMSEs of the historical expert model and the poor performance of the Naive model while the extended Expert Model has better performance when adding the covariates. In addition to this introduction, the report comprises sections on the dataset, methods, forecast performance evaluation, results, and limitations.

2 Data

The project uses data from the European Network of Transmission System Operators for Electricity platform (ENTSOE), consisting of 214 variables on electricity in France and 7 trading countries. The research model focuses on 24 variables, totaling 76,631 hourly observations from January 4, 2015, to October 2, 2023. Variable selection is based on availability during the research period and high correlation with France's day-ahead prices. The dataset is split into a training set (5 years), a validation set (1,003 days), and testing sets (365 days). The overall dataset will be divided into main 2 features as follows:

(1) Features of France:

- Day-ahead (DA) electricity prices of France. Figure 1 highlights France's DA electricity prices from 2015 to 2020 quite stable, noting a 2016 spike reaching 874.01 EUR/MWh due to nuclear plant shutdowns (Lebrouhi and Kousksou, 2022). Prices rose in mid-2021 to end-2022, influenced by the Ukraine-Russia war, impacting fuel prices and inflation (European-Commission, 2022). An outlier on April 4, 2022, when price reached almost 3000 EUR/MWh, is excluded for better prediction. The data exhibits a seasonal pattern, with winter months seeing price rises due to heating demand, while summer prices dip. Summary statistics reveal significant volatility, with a standard deviation of 94.31, fluctuating around a mean of 73.98 and a median of

43.62. The range spans a minimum of -134.94 on 02 June 2023 to 874.01 on November 7, 2016.

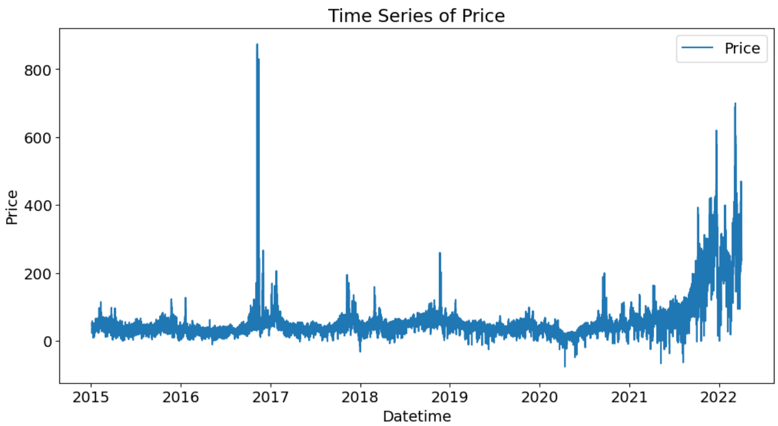

Figure 1: Day-ahead price on the training and validation set

In addition to France's day-ahead prices, we consider features related to demand and electricity generation, such as France's load day-ahead electricity; Output generation from nuclear power, hydrowater reservoirs, fossil gas, and fossil oil; Renewable energy features (wind and solar) exhibit minimal correlation with day-ahead prices, leading to their exclusion. Instead, the model focuses on "Residual Load" representing electricity demand not fulfilled by variable renewable sources, offering insights into day-ahead price movement.

(2) Interconnection features: Emphasizing cross-border exchanges with Germany, Spain, Switzerland, Italy, and Belgium, DA prices and loads from these countries are considered. Due to data limitations, features from Great Britain are excluded. Given France's significant reliance on imported fossil fuels, Oil, Gas, and Coal prices are also being considered.

Our final dataset spans 24 features, with additional lag values to enhance predictive power. Due to the auto-correlation observed for all hours, we select lag values from 1, 2, and 7 days to capture daily and weekly patterns in France's DA prices. Additional predictors include the last hour's price, and daily maximum, and minimum prices to capture extreme price variations and potential peaks during specific hours. The model comprises 33 predictors, with unlagged data from other DA prices and loads and specific lag 1-day values for electricity generation and 2 days for fuel prices.

3 Methods

In this section, we provide an overview and description of the various methods used in this project, as well as the software utilized for their implementation.

3.1 Naive Model

A simple model is given by the following model equation (3.1) for each hour s $\in$ S

$$Y_{d,s} = \begin{cases} Y_{d-7,s} + \varepsilon_{d,s} & \text{d is Mon, Sat, or Sun} \\ Y_{d-1,s} + \varepsilon_{d,s} & \text{other day d of the week} \end{cases} \tag{3.1}$$

Observing a distinct correlation between electricity prices in France and specific weekdays, weekends consistently exhibit low-price trends due to decreased demand. Mondays follow a similar pattern due to weekend residuals thus we take the price of 7 days for these days, while Tuesdays to Fridays maintain higher prices typical of working days. This unique linear model has no parameters and thus no parameter estimation is required and this serves as a benchmark for further research.

3.2 Expert Model

A linear autoregressive model is given by the following equation 3.2 for each hour s $\in$ S:

$$\begin{aligned} Y_{d,s} =& \beta_{s,0} + \beta_{s,1}Y_{d-1,s} + \beta_{s,2}Y_{d-2,s} + \beta_{s,3}Y_{d-7,s} + \\ & \beta_{s,4}DoW_d^1 + \beta_{s,5}DoW_d^6 + \beta_{s,6}DoW_d^7 + \varepsilon_{d,s} \end{aligned} \tag{3.2}$$

As observed for auto-correlation across all hours, we have autoregressive effects at lag 1 and 7 and also at lag 2. Moreover, there are three parameters $\beta_{s,4}$, $\beta_{s,5}$ and $\beta_{s,6}$ that describe the day-of-the-week effects for day-of-the-week dummies DoW_d^k specify this impact. Two extended versions of the expert model also be utilized by adding the last hours of the previous day's price which is highly correlated to the DA price of the following day and naming it as **expert.last** in equation 3.3.

$$\begin{aligned} Y_{d,s} =& \beta_{s,0} + \beta_{s,1}Y_{d-1,s} + \beta_{s,2}Y_{d-2,s} + \beta_{s,3}Y_{d-7,s} + \beta_{s,5}DoW_d^1 \\ & + \beta_{s,6}DoW_d^6 + \beta_{s,7}DoW_d^7 + \beta_{s,4}Y_{d-1,S-1} + \varepsilon_{d,s} \end{aligned} \tag{3.3}$$

The other version is the extension of the expert last model where we added the DA forecasts of discussed fundamentals and fuel prices and named it as **expert.redav** in equation 3.4

$$\begin{aligned} Y_{d,s} =& \beta_{s,0} + \beta_{s,1}Y_{d-1,s} + \beta_{s,2}Y_{d-2,s} + \beta_{s,3}Y_{d-7,s} + \beta_{s,4}Y_{d-1,S-1} + \beta_{s,5}Y_{d-1,min} \\ & + \beta_{s,6}Y_{d-1,max} + \beta_{s,7}DoW_d^1 + \beta_{s,8}DoW_d^6 + \beta_{s,9}DoW_d^7 + \beta_{s,10}X_{d,s}^{Load} \\ & + \beta_{s,11}X_{d,s}^{DARES} + \beta_{s,12}X_{d,s}^{Coal} + \beta_{s,13}X_{d,s}^{Gas} + +\beta_{s,14}X_{d,s}^{Oil} + \varepsilon_{d,s} \end{aligned} \tag{3.4}$$

3.3 LASSO and Ridge regression

LASSO, short for Least Absolute Shrinkage and Selection Operator, and Ridge Regression are statistical techniques addressing overfitting and multicollinearity of predictors in linear regression. These techniques are called shrinkage since they add a penalty term called the "ℓ_1-regularization" (for LASSO in equation 3.5) and "ℓ_2 regularization" (for Ridge Regression in equation 3.6) (Streib and Dehmer, 2019)

$$\hat{\hat{\beta}}_{\lambda,s}^{\text{lasso}} = \underset{\beta \epsilon \mathbb{R}^D}{\text{arg min}} \|\tilde{Y}_s - \tilde{X}_s\beta\|_2^2 + \lambda_s\|\beta\|_1 \tag{3.5}$$

$$\hat{\hat{\beta}}_{\lambda,s}^{\text{ridge}} = \underset{\beta \epsilon \mathbb{R}^D}{\text{arg min}} \|\tilde{Y}_s - \tilde{X}_s\beta\|_2^2 + \lambda_s\|\beta|_2^2 \tag{3.6}$$

$\hat{\hat{\beta}}_{\lambda,s}^{\text{lasso}}$ and $\hat{\hat{\beta}}_{\lambda,s}^{\text{ridge}}$ are obtained by minimizing the loss function, which contains the residual sum of squares term $\|Y_s - X_s\beta\|_2^2$ and the penalty terms. For the penalty, LASSO adds the sum of the absolute values of the coefficients ($\|\beta\|_1 = \sum_{k=1}^K |\beta_k|$) and Ridge regression adds the sum of the squared values of the coefficients (($\|\beta\|_2^2 = \sum_{k=1}^K |\beta_k|^2$) to the loss function. Parameter λ controls the strength of the ℓ_1 and ℓ_2 regularization for each hour s $\in$ S. A higher λ increases the shrinkage of coefficients, forcing more coefficients to become exactly zero (for LASSO) or reduce their magnitude but not becoming exactly zero (for Ridge regression). This helps prevent overfitting and balance between model complexity and generalization performance. Otherwise, due to the ability of the shrinkage parameter to exactly 0, LASSO can help the model "get rid of" irrelevant variables, resulting in a simpler and easier-to-interpret model. For that reason, LASSO is also useful for variable selection.

3.4 Elastic Net

Elastic Net is a regularization technique used in linear regression and related models. It combines both ℓ_1 (LASSO) and ℓ_2 (Ridge) regularization penalties to address the limitations of each method. Elastic Net is representing a trade-off between the ℓ_1-norm (LASSO) and the ℓ_2-norm (Ridge) by the combination in the penalty term $\mathcal{P}_\alpha(\beta) = \alpha(\lambda_{s,1}|\beta|_1 + (1-\lambda_{s,2})\|\beta\|_2^2)$ which lead to the equation for Elastic Net as in equation 3.7 (Streib and Dehmer, 2019). This balance is pivotal in achieving an effective compromise between variable selection (sparsity) and regularization (shrinkage).

$$\hat{\hat{\beta}}^{\text{elasticnet}}_{\Omega_{\lambda_s},s} = \underset{\beta \in \mathbb{R}^D}{\arg\min} \|\tilde{Y}_s - \tilde{X}_s\beta\|_2^2 + \lambda\mathcal{P}_\alpha(\beta) \tag{3.7}$$

with α is a mixing ratio, controls the balance between the ℓ_1 and ℓ_2 penalties. $\lambda_{s,1}$ and $\lambda_{s,2}$ are regularization parameters for the ℓ_1-norm and ℓ_2-norm penalties for each hour s $\in$ S.

To ensure the effectiveness of LASSO, Ridge regression, and Elastic Net, all input variables must share the same unit of measurement. Therefore, scaling is implemented in this research before applying regularization techniques. The scaling process involves standardizing each feature in the regression matrix (X) and the response variable (Y) by subtracting the mean and dividing by the standard deviation. This step ensures uniform application of regularization penalties across features, preventing bias towards variables with larger scales and facilitating optimal results in regularization.

3.5 Root Mean Squared Error

To evaluate the accuracy of forecasts in this project, we employed Root Mean Squared Error (RMSE). RMSE is computed by taking the root of the average of the squared differences between forecasted values and actual values over a specified period as the formula: $RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{s,t} - \hat{y}_{s,t})^2}$. RMSE is the key metric for evaluating the performance of forecasting methods or models and helps in comparing different forecasting techniques. A high RMSE implies that the fitted values are further away from the actual data, the lower RMSE indicates better forecasting accuracy.

3.6 Diebold-Mariano test

To have more insight into the statistical significance of forecast accuracy, the Diebold-Mariano test is utilized. To compare the forecast accuracy of two models A and B, the key objective is to compare a loss differential of two losses $L_{A,t}$, measures the loss of model A at time t, and $L_{B,t}$, measures the loss of model B at a certain time t, with loss can be the absolute error loss ($L_{A,t} = \|e_{At}\|$) or the squared-error loss ($L_{A,t} = \|e_{At}\|^2$). The series $\Delta_{A;B;t} = L_{A,t} - L_{B,t}$ measures the difference between both losses. The null hypothesis will be $H_0 : E(\Delta_{A;B;t}) = 0$ with all t, which means two forecasts A and B have equal forecast accuracy, versus the alternative hypothesis $H_1 : E(\Delta_{A;B;t}) \neq 0$, which means two forecasts have different levels of accuracy. The test works in analogy to a t-test or z-test, where we can compare the differences between two means. Under the null hypothesis, the test statistics DM is asymptotically $N(0,1)$ distributed. The null hypothesis of no difference will be rejected if the computed DM statistic falls outside the range of -$z_{\frac{\alpha}{2}}$ to $z_{\frac{\alpha}{2}}$ (Raj, 2020)

4 Results

This section is dedicated to analyzing the data using the methods defined in the preceding section. It is divided into the following subsections:

4.1 Hyperparameter

λ value is selected during model tuning through 5-fold cross-validation using $LassoCV$, $RidgeCV$, and $ElasticNetCV$ functions in Python. This process automatically searches for the optimal λ for each hour, minimizing prediction error on a validation set. Coordinate descent estimates the lasso or ridge estimator on a grid of λ values, $\Lambda = 2^{-5}, ..., 2^{0.5}$ over 100 steps. For elastic net, the α ratio is set at 0.5, achieving a balanced combination of sparsity-inducing LASSO and shrinkage-inducing Ridge penalties. The best coefficients aligned with the choice of λ also are recorded for all hours for all variables in matrix form for future forecasting applications, however, due to the large dimension of the results, the coefficients will not be mentioned in this report. Otherwise, the shrinkage behavior of β value when increasing λ also be observed. Figure 2 shows the shrinkage behavior at hour 9 for LASSO and Ridge regression.

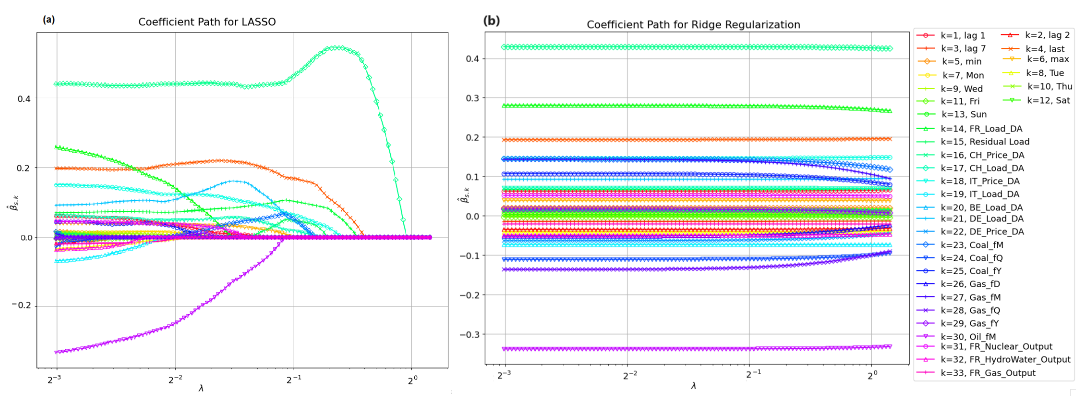

Figure 2: Shrinkage behavior of LASSO and Ridge regression.

LASSO systematically shrinks less significant parameters toward zero, favoring variable selection. At λ=0.1, about 80% of coefficients become zero, aligning with intentionally sparse data. Figure 2 (a) highlights six predictors with slow-decaying coefficients, including features related to Switzerland's, Italy's and Germany's load, oil prices, local load forecasts, and the preceding day's last-hour prices. Unlike LASSO, Ridge regression shrinks coefficients toward zero without setting them precisely to zero. Increasing the regularization parameter (λ) in Ridge intensifies the shrinkage effect. Figure 2 (b) shows that as λ increases, coefficients approach smaller values. Ridge regression's power lies in reducing coefficients to alleviate collinearity between predictors. Elastic Net exhibits similar shrinkage behavior to LASSO, with coefficients gradually approaching zero later, attributed to Elastic Net's inclusion of the Ridge penalty component.

4.2 Model Performance

RMSEs are calculated to assess model performance on both the evaluation and testing sets, providing a basis for comparing predictive accuracy. RMSE results reveal Elastic Net's supremacy in the validation set with the lowest RMSE of 22.12, followed closely by LASSO (23.73), and Ridge regression trailing at 25.13. Elastic Net's success stems from effectively blending LASSO's feature elimination and Ridge regression's coefficient reduction for an accurate and concise model. In contrast, historical expert models, including Naive (45.39) and expert.last model having the highest RMSEs, highlighting the limitations of solely relying on past data. Models like expert.redav surpass historical ones but fall short of regularization models (Figure 3(a)). On the testing set, RMSEs are considerably smaller. Ridge regression unexpectedly outperforms others at RMSE of 5.44, followed by LASSO (5.89), and Elastic Net (6.069). This underscores Ridge regression's practical application. Similar to the validation set, all historical models have poorer performance compared to LASSO-based models.

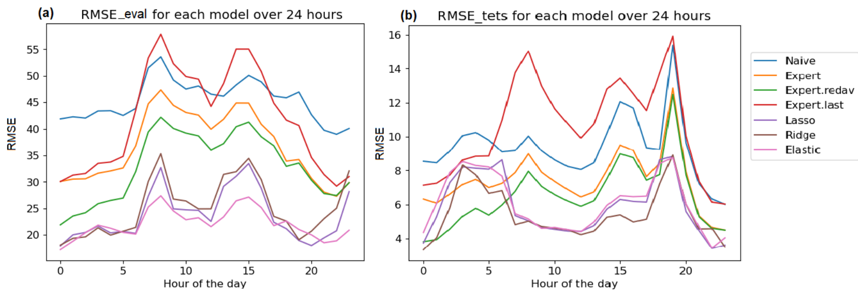

Figure 3: RMSEs by different models for different hours.

Moreover, a comprehensive evaluation of model performance across all hours for both valuation and testing sets are depicted in Figure 3. Notably, LASSO, Ridge, and Elastic Net exhibit closely aligned and nearly identical performances throughout all hours. Conversely, the historical models display poorer performance, with higher RMSEs across various hours. An exception is observed from 1 to 6 am in the testing set ((Figure 3 (b)), where the expert.redav model achieves the smallest RMSEs. Notably, all models face challenges in accurately forecasting day-ahead prices during specific hours, such as 8 in the morning and hours 15 in the afternoon. The off-peak hours (from 0 to 5 and around hour 20) consistently yield the smallest RMSEs.

4.3 Diebold-Mariano test results

In addition to the RMSE results, we also conducted DM tests across various models for both the validation set and the testing set. The outcome is visually depicted through a heat map in Figure 4. These maps illustrate the p-values associated with the null hypothesis of the forecasts. The low p-values (green color) show statistical significance that the model along the y-axis is significantly more accurate than the forecast along the x-axis. On the other hand, shades closer to red (p-values approaching 0.1) suggest weaker evidence against the null hypothesis.

In both validation and testing sets, the naive and expert.last models exhibit inferior performance compared to others. In the validation set, LASSO-based models outperform naive and expert.last. Elastic Net notably outperforms Ridge regression but there's no compelling evidence of superiority over expert and expert.redav.

On the testing set, all historical models show lower forecast accuracy than LASSO-based models. Overall, there is no statistical evidence to conclusively establish superior performance within the LASSO-based models only, Ridge regression slightly outperforms LASSO, as suggested by the data.

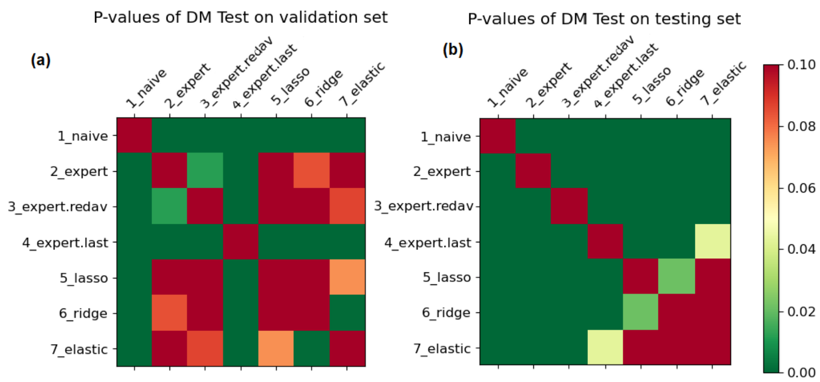

Figure 4: RMSEs by different models for different hours.

5 Conclusion

In summary, this research addresses challenges in forecasting day-ahead electricity prices in the dynamic energy context of France. Focused on high-dimensional projections, we employ three shrinkage methods: Ridge regression, LASSO, and Elastic Net and four benchmark models: Naive model, Expert model, Expert extend the last hour and Expert extend with covariates. The analysis reveals Elastic Net, with the combination of LASSO and Ridge regression, outperforms other models in the validation set but in practical scenarios Ridge regression stands out as the best performer, as seen in low RMSEs. LASSO and Elastic Net exhibit similar shrinkage behaviors, while Ridge regression minimizes collinearity. In contrast, all benchmark model has high RMSE, especially the naive and expert.last in both the validation and testing sets. The expert.redav has better performance than the other benchmark model and even better than the LASSO-based model in a specific time. In general, all models face a challenge in forecasting the peak hours (around hour 8, hour 14, and 15) and have good performance in the off-peak hours. The Diebold-Mariano test reveals subtle differences in forecasting accuracy among LASSO-based models, with Elastic Net outperforming Ridge regression in sample forecasting and Ridge regression slightly edging out LASSO in out-of-sample forecasting. The study underscores the importance of incorporating covariates in forecasting dynamic price changes, advocating for more complex models. Additionally, it emphasizes the efficacy of Ridge regression in real-world scenarios.

Bibliography

European-Commission. *State aid: Commission consults Member States on proposal to amend Temporary Crisis Framework*, 2022.

Schall Eric Lamrani Bilal Yassine Chaibi Lebrouhi, Badr Eddine and Tarik Kousksou. *Energy Transition in France.*, 2022. URL https://www.mdpi.com/2071-1050/14/10/5818.

Ashwin Raj. *A Quick Guide to Random Forest Regression*, 2020.

Frank Streib and Matthias Dehmer. *High-Dimensional LASSO-Based ComputationalRegression Models: Regularization, Shrinkage,and Selection*, 2019.