Authors: Ly Le Thi
Supervisors: Dr. Paul Navas Alban
Prof. Dr. Matei Demetrescu
Prof. Dr. Carsten Jentsch

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

1- Motivation

- ✓ Accurate day-ahead electricity price forecasting is crucial for grid efficiency and reliability, benefiting stakeholders
- ✓ Challenges from the dynamic and complex energy landscape (integration of energy sources and cross-border electricity trading) and recent geopolitical events.
- ✓ Inspired by the paper of (Babii and Striaukas, 2023), focus on the high-dimensional projections advancement for time series forecasting
- ✓ 3 Shrinkage methods (Ridge Regression, LASSO, and Elastic Net) are employed
- ✓ Two benchmark models are utilized: Naive model, Expert model

2. Data – Regressors selection

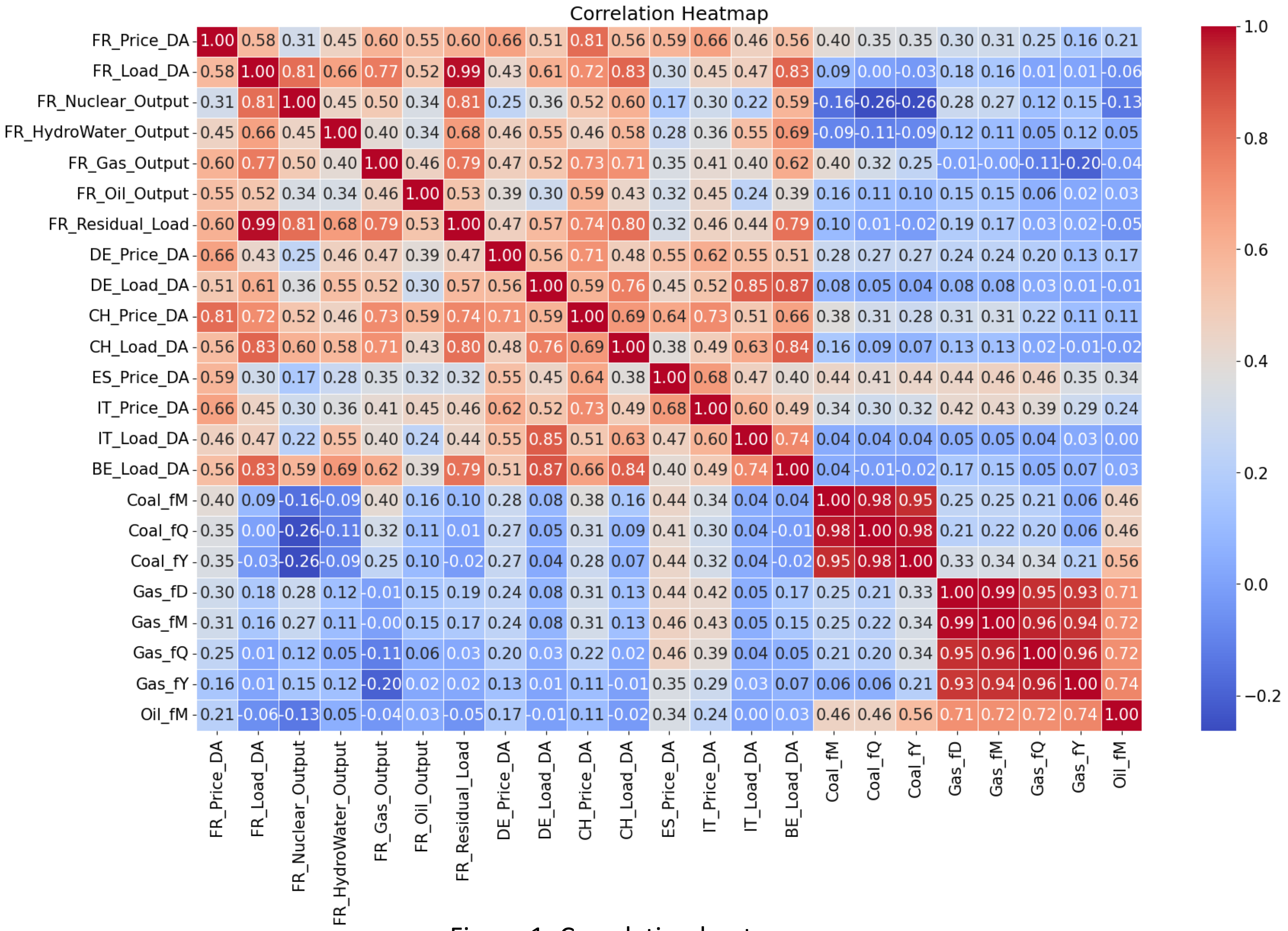


Figure 1: Correlation heatmap

Two main features:

- Feature of France: Price, Load, Output, Residual Load

- Interconection features: Price, Load of other countries + Oil, Gas and Coal prices

- Total of 33 variables: including 24 variables and its lagged value

- Data taken from ENTSOE

2. Data – Regressors selection

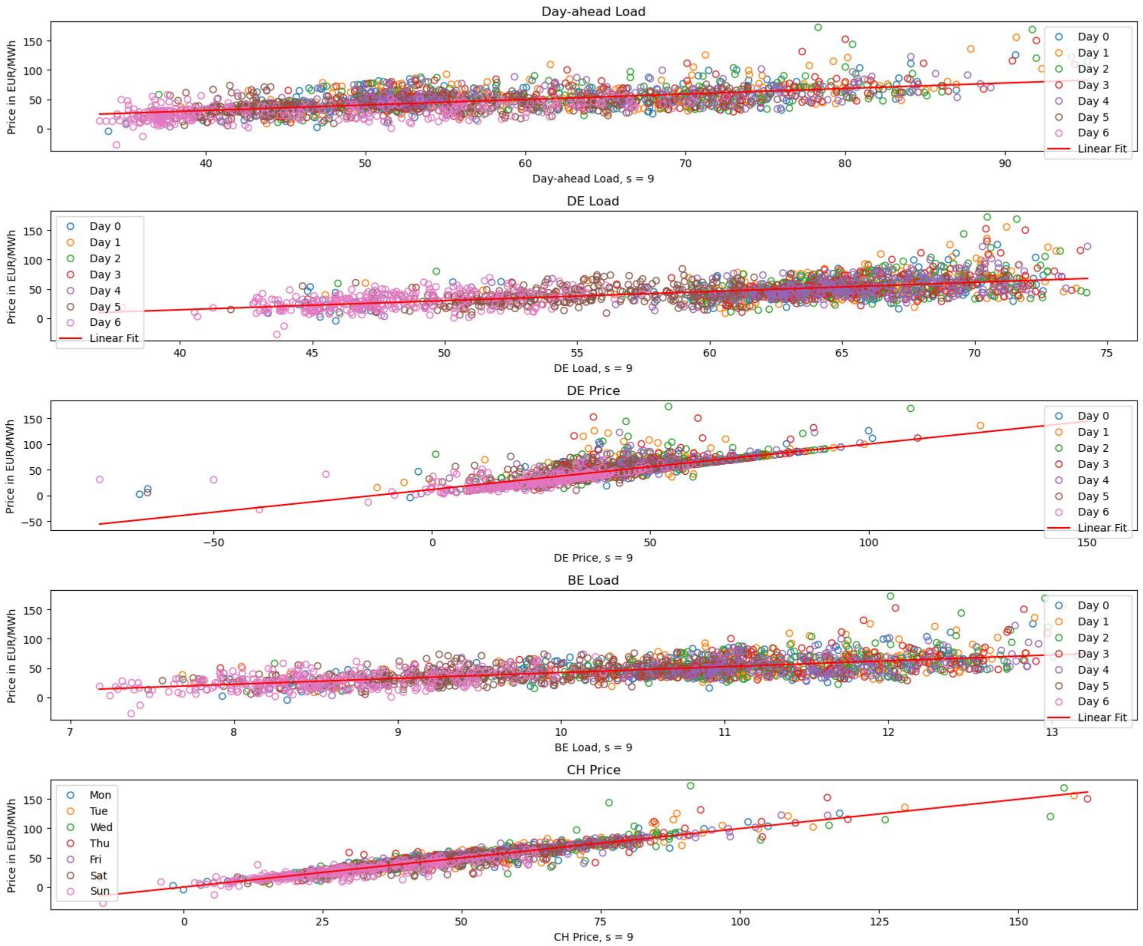


Figure 2: Relationship between some variables and DA price of France

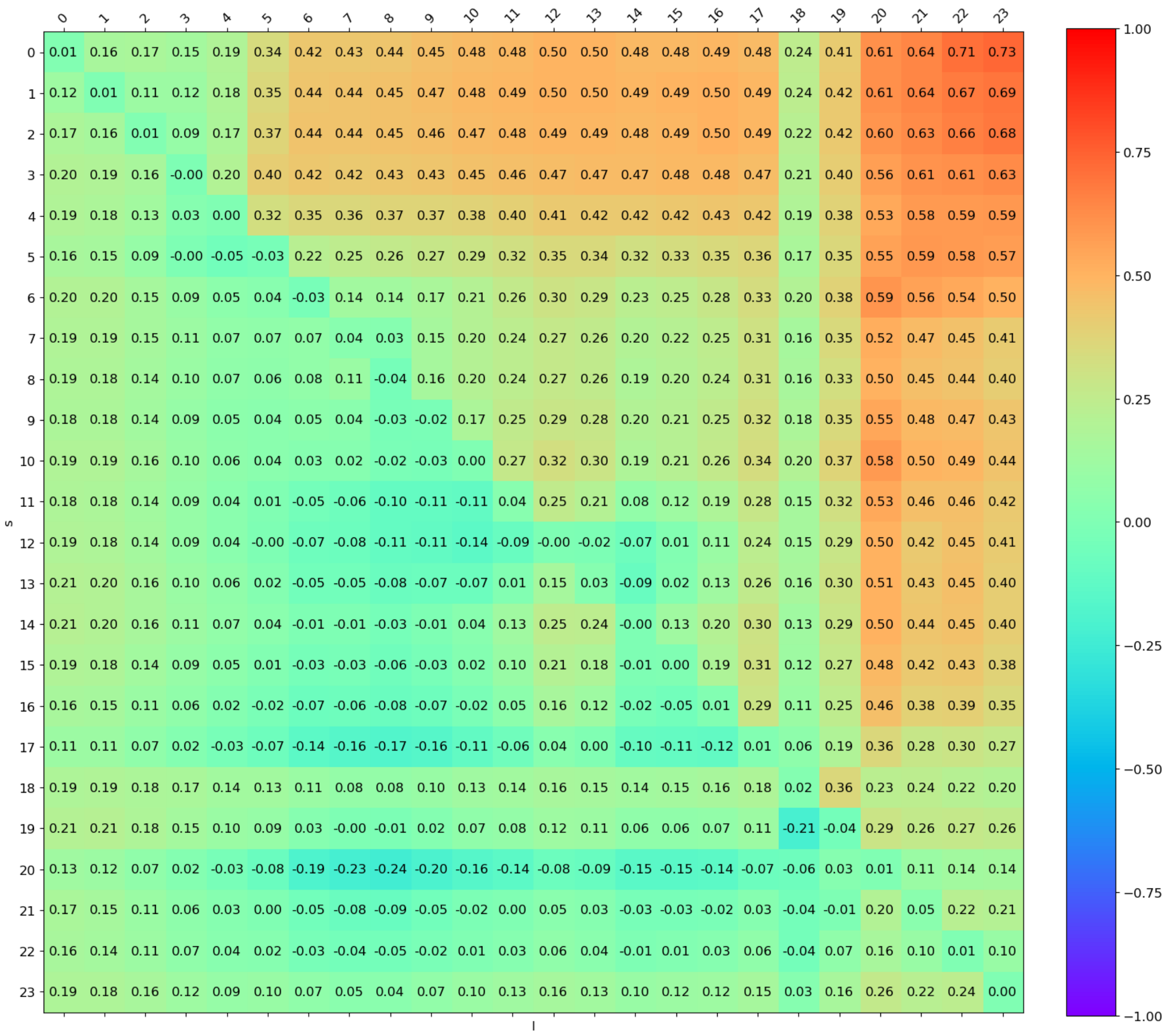


Figure 3: Correlation heatmap for all hours

2. Data - Target variable

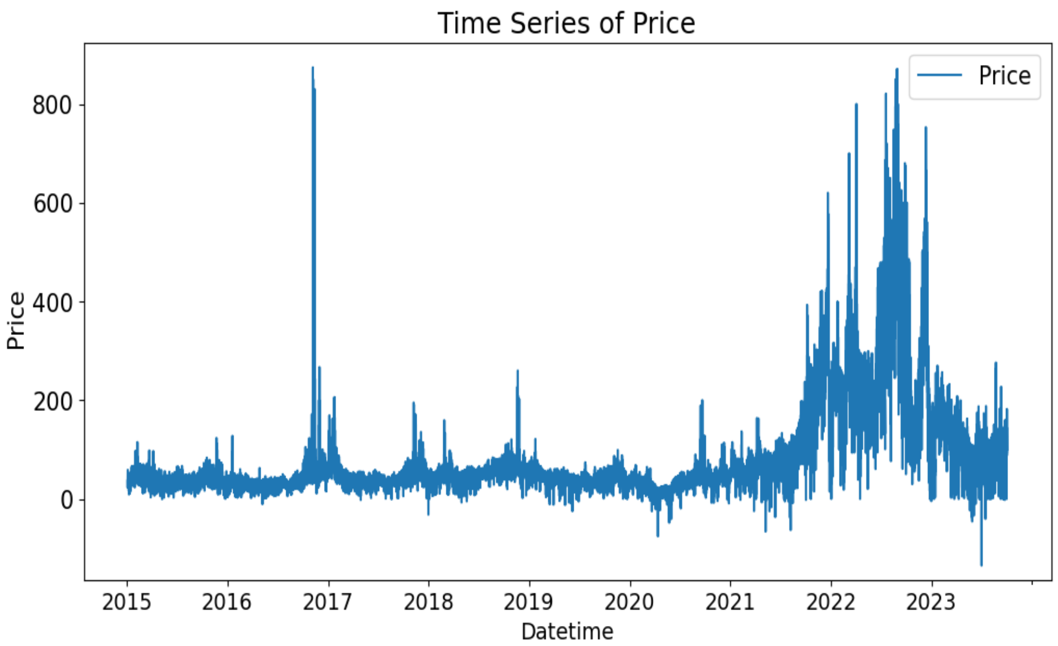


Figure 4a: France's Day-ahead price over time (after outlier cleaning on 2022-04-04)

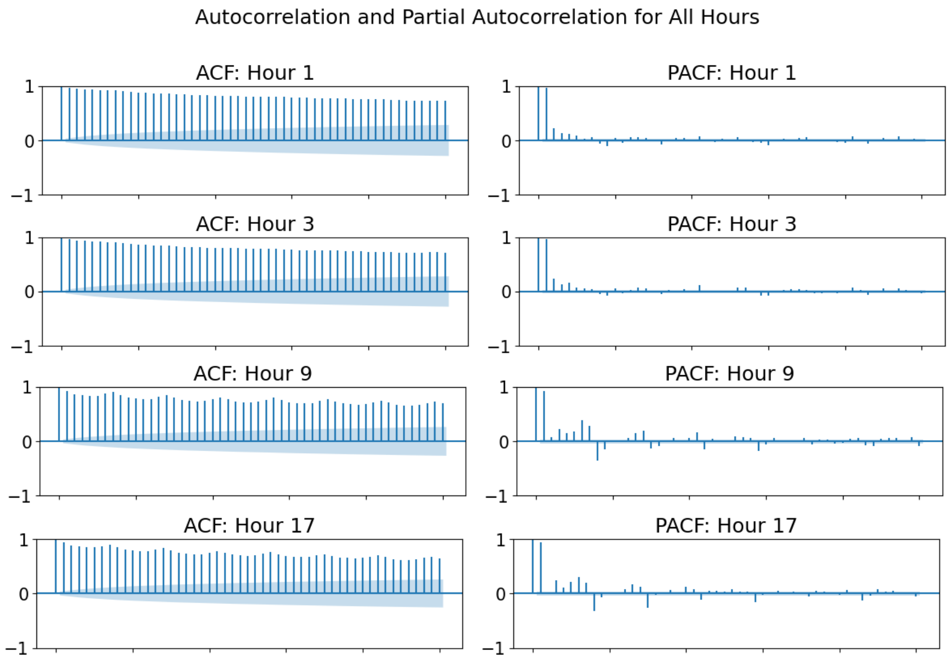
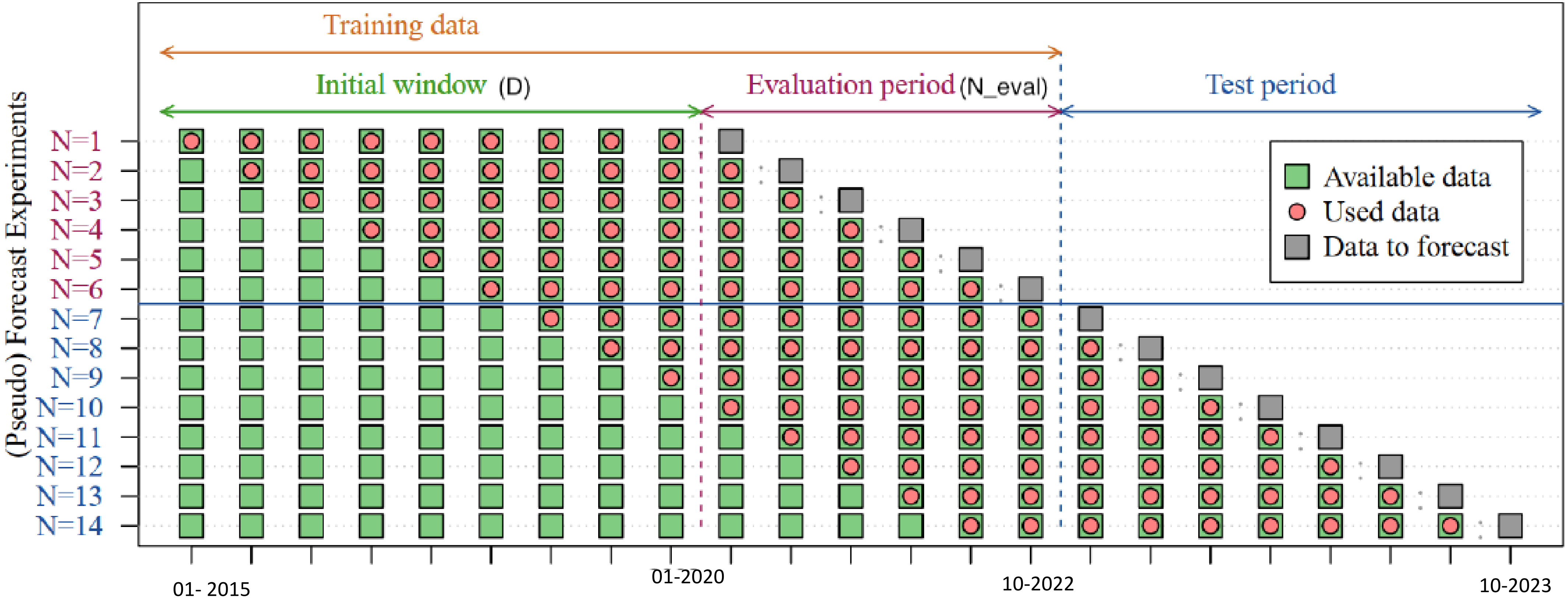


Figure 4b: Auto-correlation structure

Table 1: Summary statistics of the prices in France

Prices	Mean	Median	Standard Deviation	Minimum	Max
Statistics	80.41	46.52	95.26	-134.94	874.01

2. Data - Study Design



Spliting data and rolling forecast technique (Sources: Hansika Hewamalage, 2023)

- Total dataset: 76631 data points of 24 variables and ist lagged values, hourly observations
- Data range: more than 8 years (from 2015-01-04 at 23:00:00 to 2023-10-02 at 21:00:00)
- Training set D= 5 years, validation set: N_eval= 2 years, and testing set =1 year

3. Methods – Benchmark models

Naive Model

$$Y_{d,s} = \begin{cases} Y_{d-7,s} + \varepsilon_{d,s} & \text{, d is Monday, Saturday, or Sunday} \\ Y_{d-1,s} + \varepsilon_{d,s} & \text{, other day d of the week} \end{cases}$$

Expert Model

$$Y_{d,s} = \beta_{s,0} + \beta_{s,1} Y_{d-1,s} + \beta_{s,2} Y_{d-2,s} + \beta_{s,3} Y_{d-7,s} + \beta_{s,4} DoW_d^1 + \beta_{s,5} DoW_d^6 + \beta_{s,6} DoW_d^7 + \varepsilon_{d,s}$$

With $Y_{d,s}$ is the DA electricity price, DoW_d^k is day-of-the-week dummies, k=1 for Monday, and k=7 for Sunday

3. Methods – LASSO-based models

fakultät statistik

LASSO

- Least Absolute Shrinkage and Selection Operator
- L1-regularization with L1-norm penalty term
- higher λ encourages sparsity, forcing more coefficients to become exactly zero
- Prevent overfitting and “get rid of” irrelevant variables => useful for variable selection

Ridge Regression

- Tikhonov regularization or L2 regularization
- Penalty term is as the ”L2-norm” or Euclidean norm
- Shrinking the coefficients by reducing their magnitude but not becoming exactly zero

Elastic Net

- A combination of the L1-norm (LASSO) and the L2-norm (Ridge) in the penalty term
- Effective compromise between variable selection (sparsity) and regularization (shrinkage)

3. Methods – LASSO-based models

$$\hat{\beta}^{\text{lasso}}_{\lambda,s} = \underset{\beta \epsilon \mathbb{R}^D}{\arg\min} \|Y_s - X_s\beta\|_2^2 + \lambda_s\|\beta\|_1$$

$$\hat{\beta}^{\text{elasticnet}}_{\Omega_{\lambda_s},s} = \underset{\beta \in \mathbb{R}^D}{\arg\min} \|Y_s - X_s\beta\|_2^2 + \lambda_s \mathcal{P}_\alpha(\beta)$$

$$\mathcal{P}_\alpha(\beta) = \alpha\|\beta\|_2^2 + (1-\alpha)\|\beta\|_1$$

Ridge Regression

LASSO

Elastic Net

$$\hat{\beta}^{\text{ridge}}_{\lambda,s} = \underset{\beta \epsilon \mathbb{R}^D}{\arg\min} \|Y_s - X_s\beta\|_2^2 + \lambda_s\|\beta|_2^2$$

- β is obtained by minimizing the loss function (SSR +penalty term)
- Parameter λ controls the strength of the L1 or L2 regularization for each hour s ∈ S
- α is a mixing ratio which controls the balance of penalty

fakultät statistik

3. Evaluation Methods

Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{s,t}-\hat{y}_{s,t})^2}$$

Diebold-Mariano test

For comparing the forecast accuracy of two forecast model A and B
Consider the difference loss function

$$\Delta_{A;B;t} = L_{A,t} - L_{B,t}$$

The null hypothesis

$$H_0 : E(\Delta_{A;B;t}) = 0 \quad \text{vs.} \quad H_1 : E(\Delta_{A;B;t}) \neq 0$$

The null hypothesis will be rejected if the computed DM statistic falls outside the range of $-z_{\frac{\alpha}{2}}$ to $z_{\frac{\alpha}{2}}$

4. Results - Hyperparameter

- Hyperparameter range: $\Lambda = 2^{-10}, \ldots, 2^{2}$ over 100 steps, tunning and selected using 5-fold cross-validation
- α ratio of Elastic Net is set at 0.5
- Best λ value for Lasso around 0.00097, for Ridge regression around 4, and Elastic Net model around 0.000976

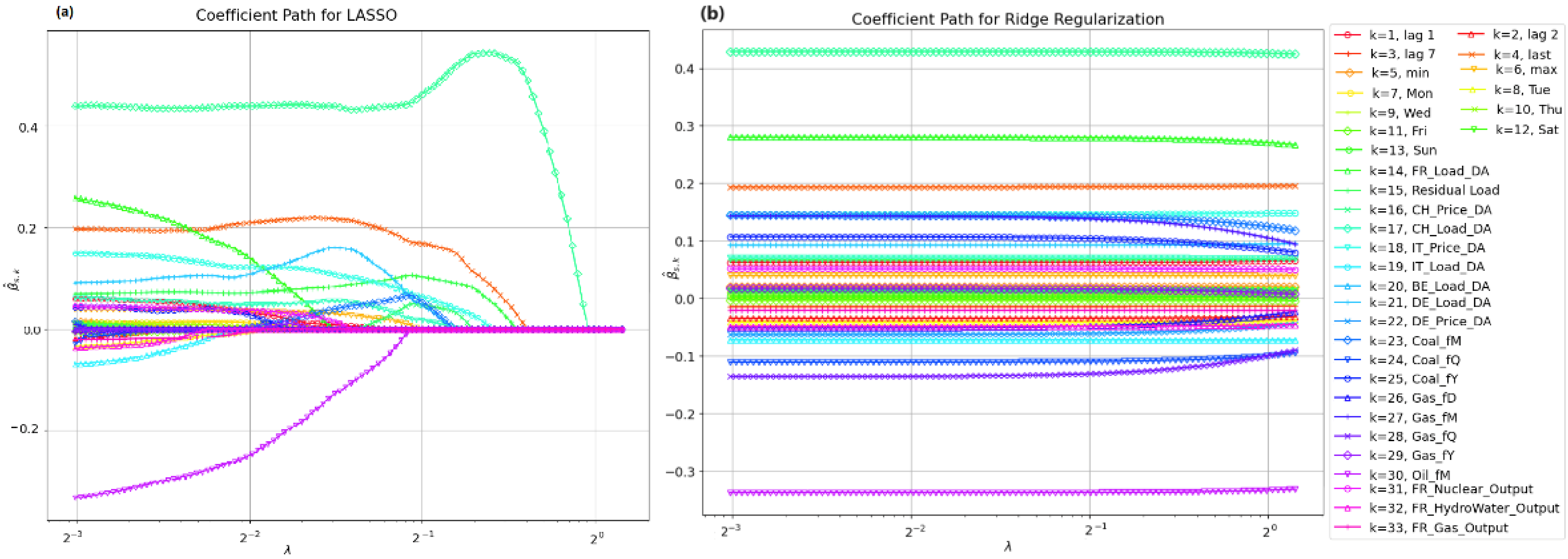


Figure 5: Shrinkage behavior of LASSO and Ridge regression.

Most important predictors: Switzerland's Load DA, price last hour of the day before, Oil price fm, Italy's Load DA, and France's Load DA and Germany's Load DA

4- Results - Model Performance

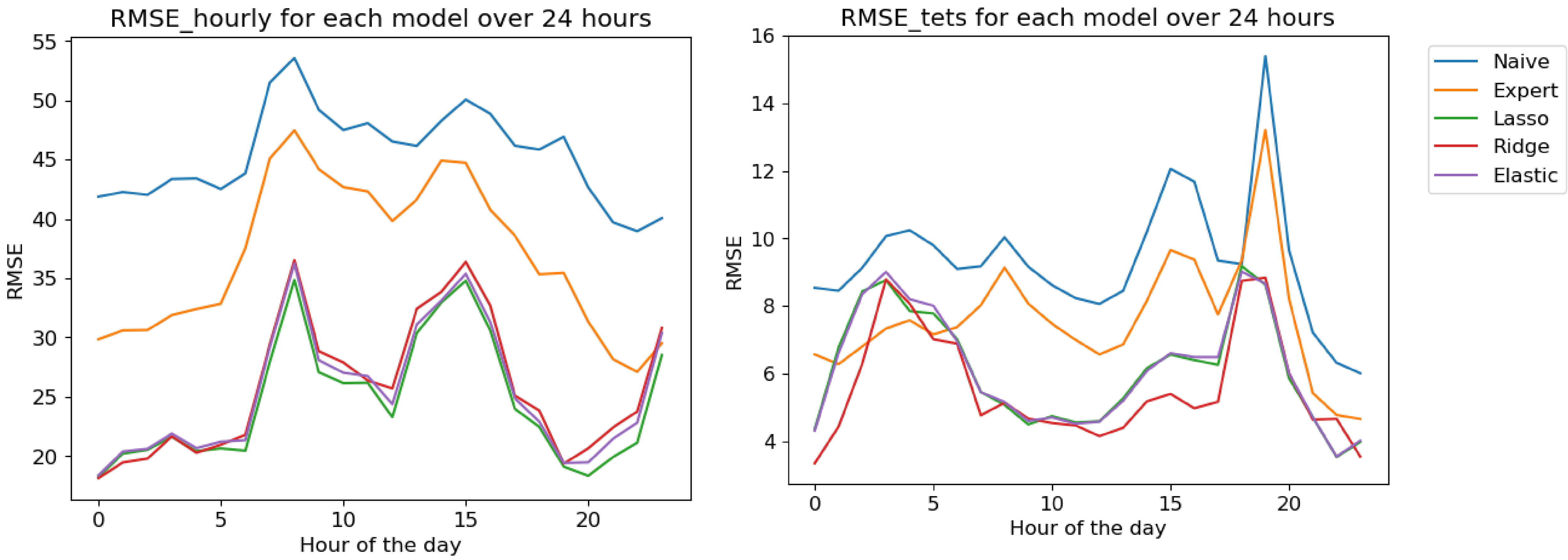


Figure 6: RMSEs by different models for different hours.

4. Results - Diebold-Mariano test

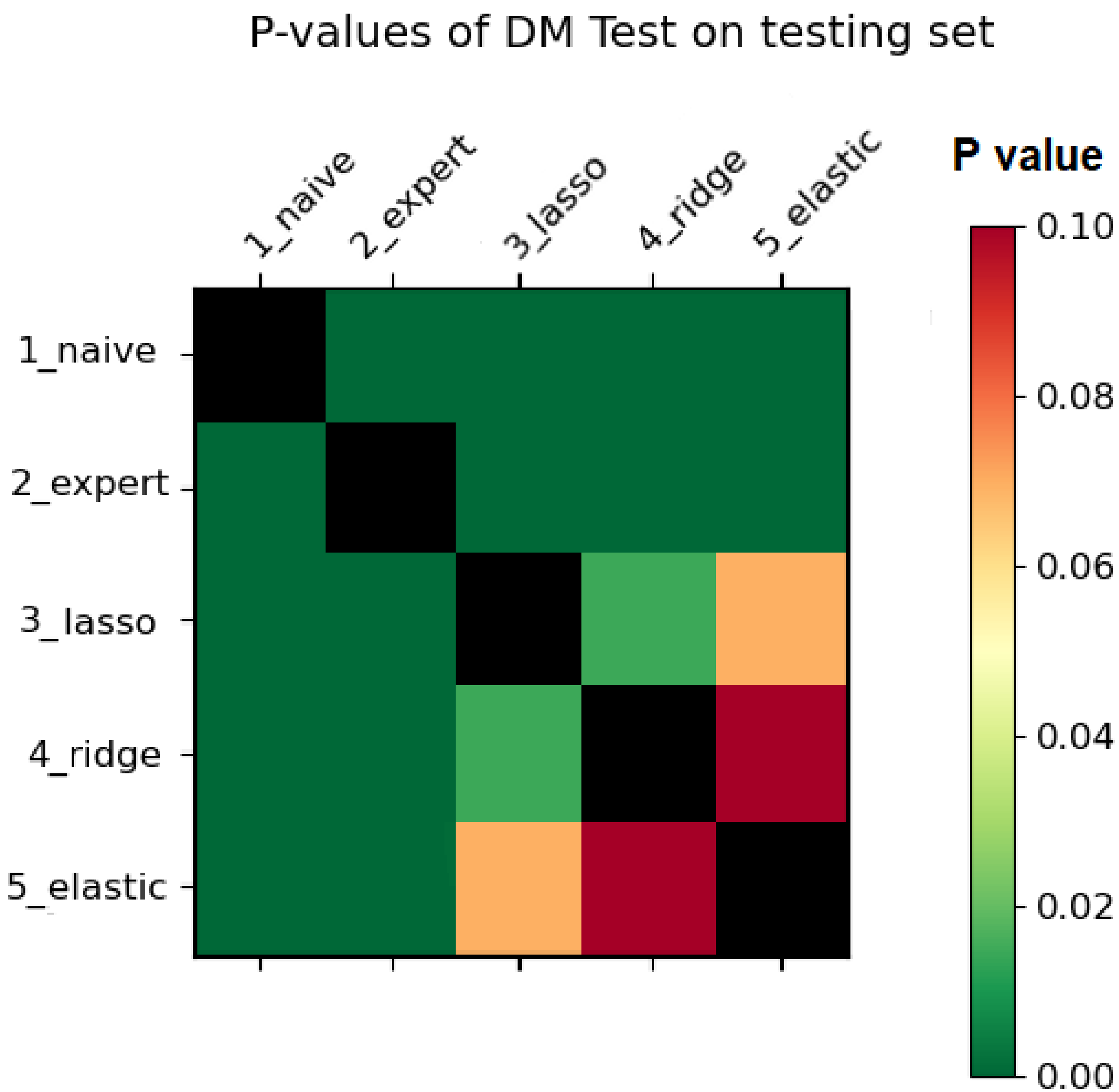
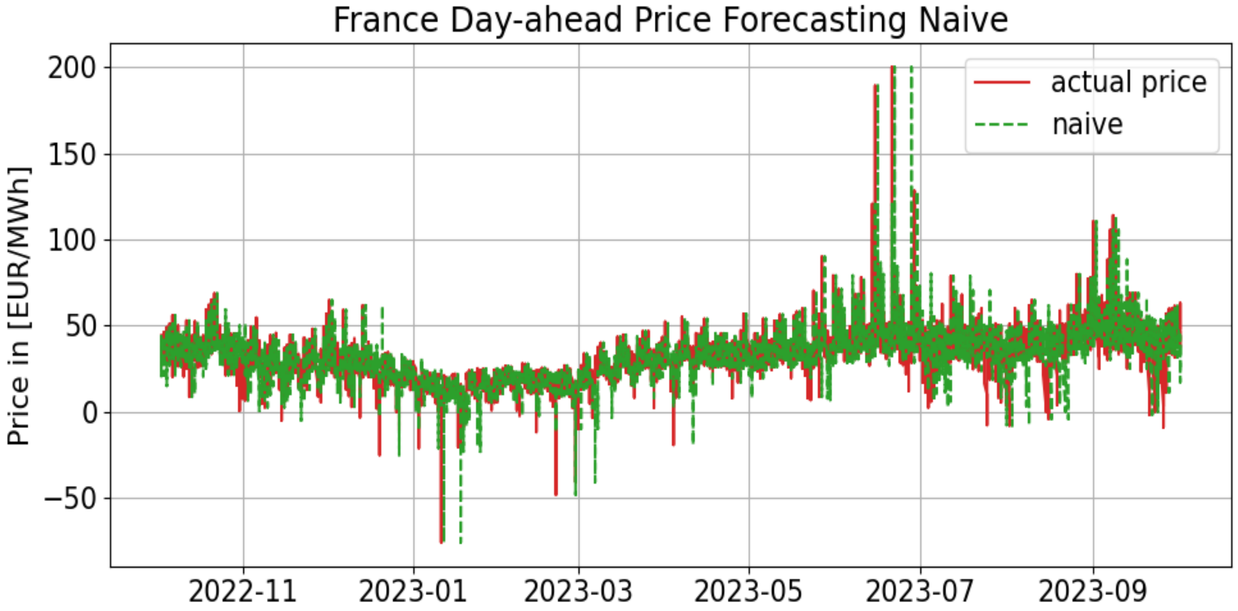
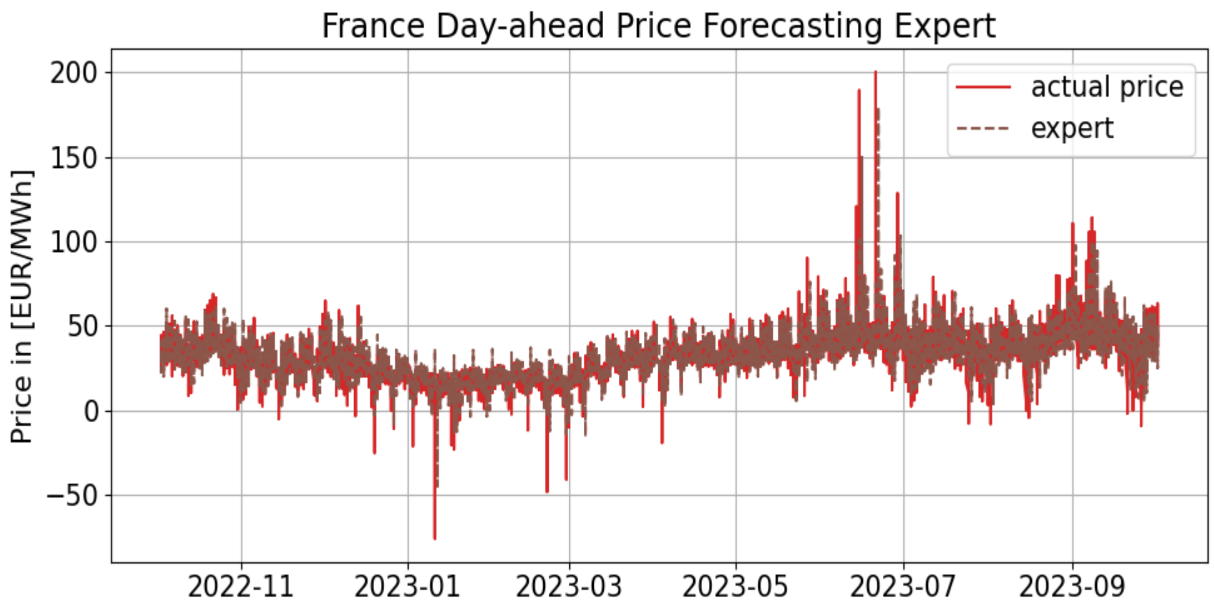
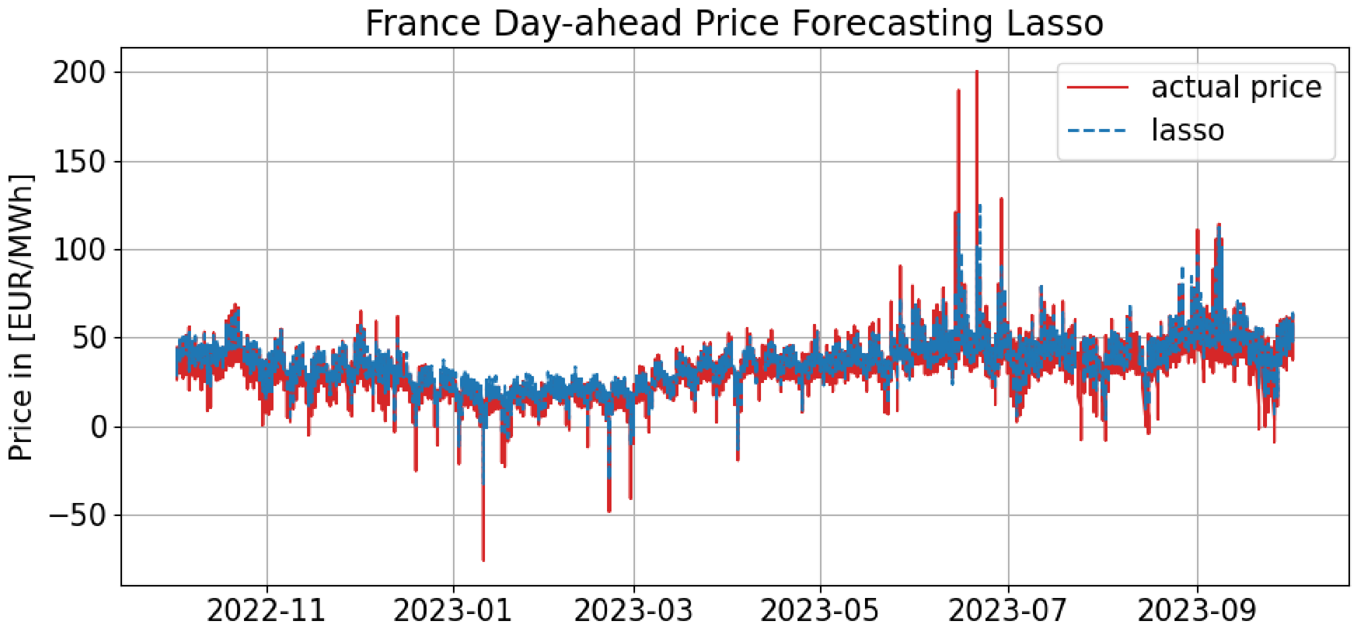


Figure 7: DM test results

4. Results – Forecast visualization

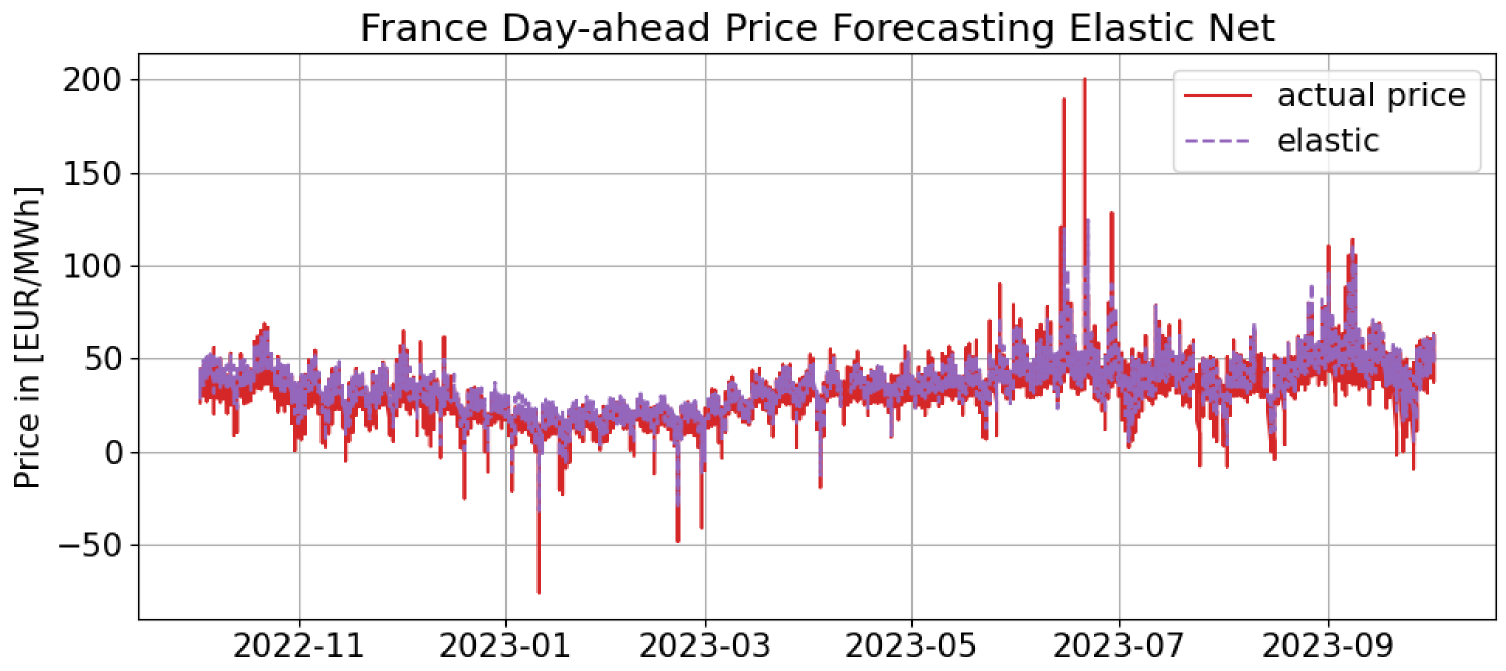


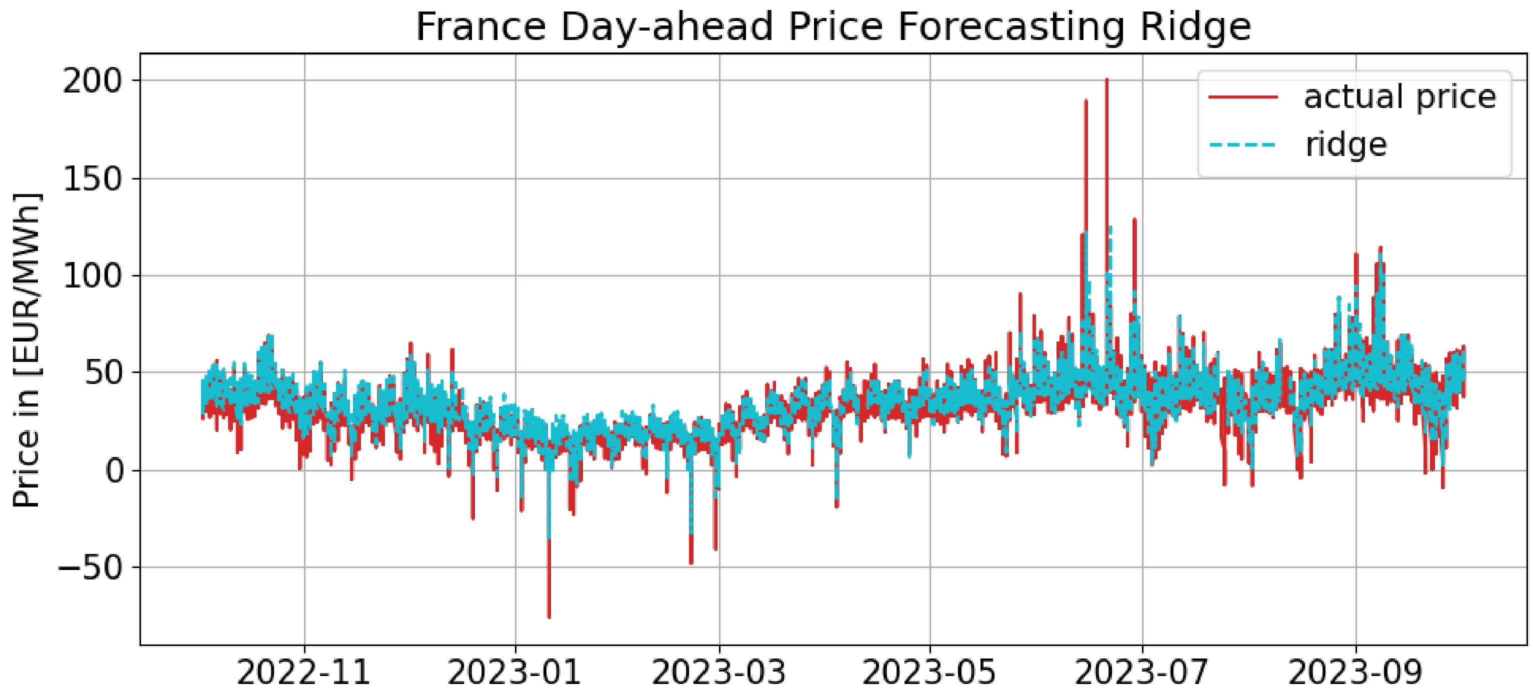




4. Results

Forecast visualization





5 - Conclusions

- LASSO-based models have close RMSE results
- LASSO outperforms other models in the validation set.
- Ridge regression stands out as the best performer in out-of-sample testing set
- The naive and expert has the worst performance, the need for more sophisticated model with other predictors than only use the historical DA price
- The Diebold-Mariano test emphasizes the forecast power of Ridge regression in practical scenarios
- The study underscores the importance of incorporating covariates and the advantage of LASSO-based models in forecasting

Thank You

6 - Bibliography

- Eric Ghysels Andrii Babii and Jonas Striaukas. Econometrics of Machine Learning Methods in Economic Forecasting, 2023
- 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 Computational Regression Models: Regularization, Shrinkage,and Selection, 2019.