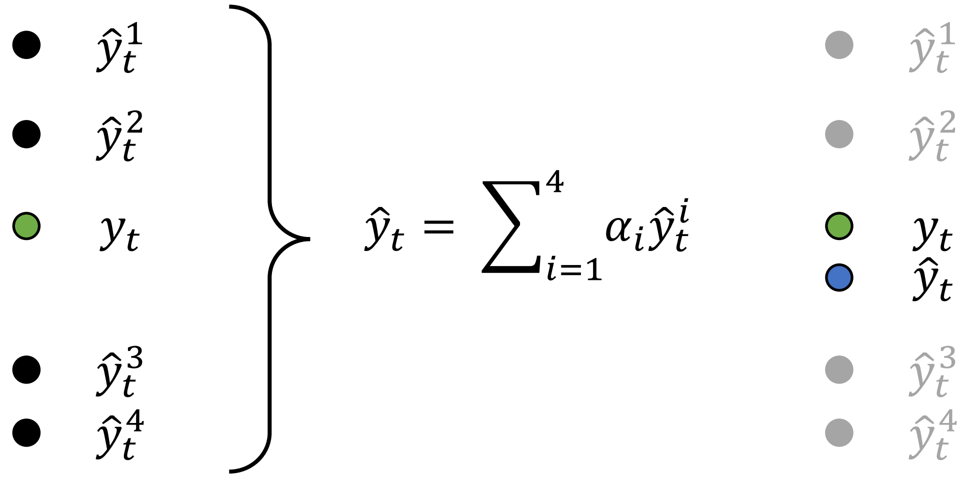
Forecasting electricity demand – ensemble methods

Jérôme Emery

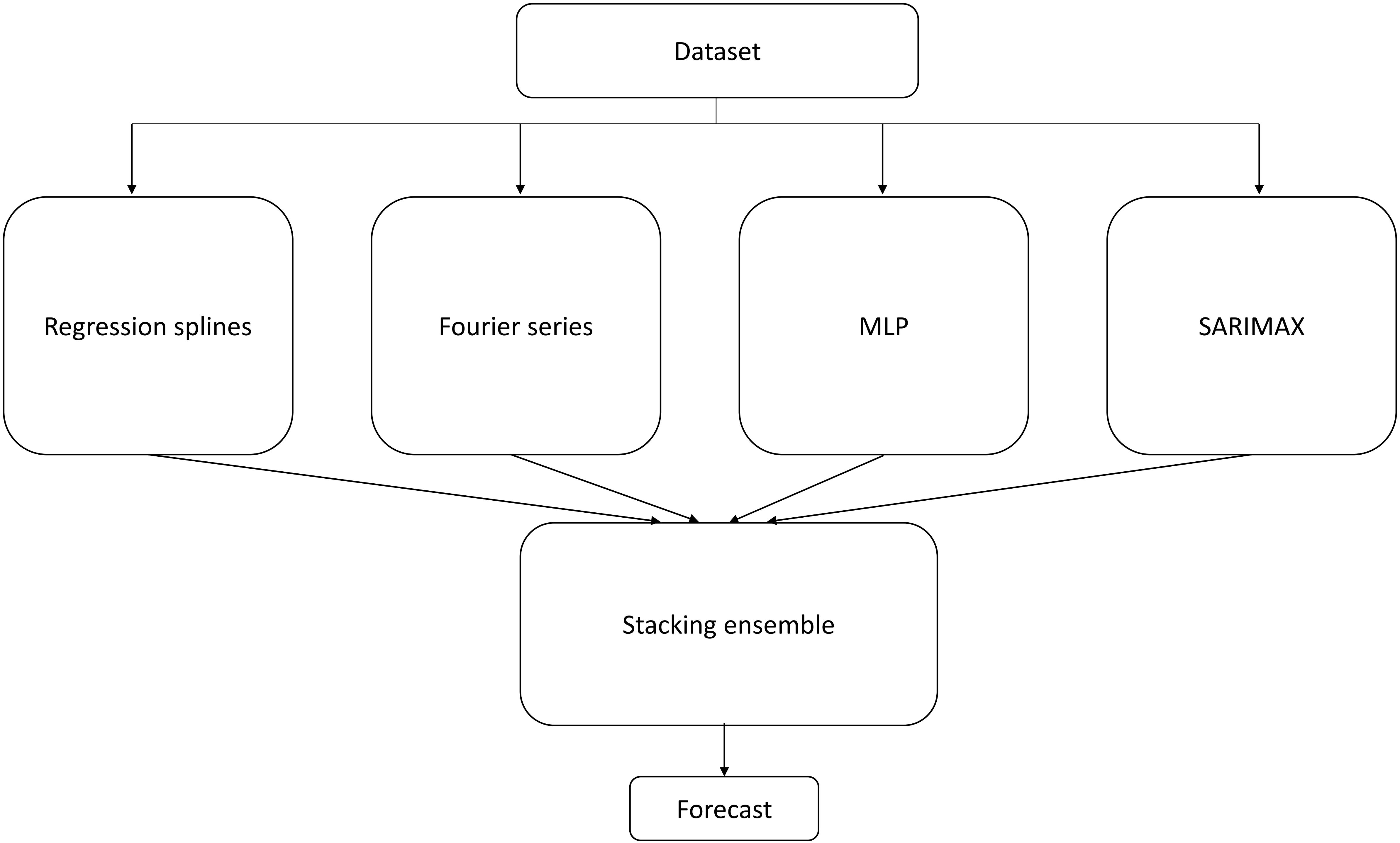
Update august 18

Ensemble methods - stacking

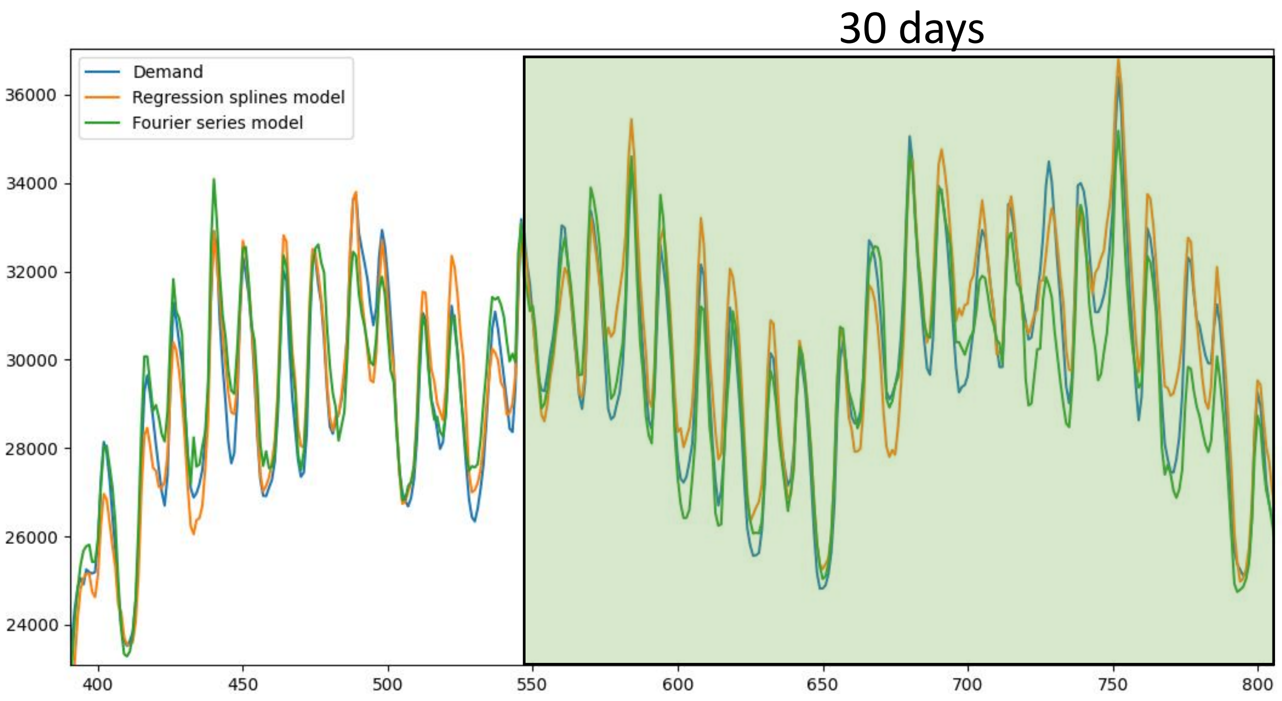
- Create many base models, average their predictions.



- Improve accuracy, reduce errors.
- How to determine weights for each forecast?


Dataset
Regression splines
Fourier series
MLP
SARIMAX
Stacking ensemble
Forecast

Rolling window linear regression



$$\hat{y}_t = c + \sum_{i=1}^{4} \alpha_i \hat{y}_t^i$$

- With c and α_i determined by least squares regression

Least squares regression

Linear regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \dots \beta_k x_k + \epsilon$

Matrix form: $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$

Model: $y = X\beta + \epsilon$

Solution: $\hat{\beta}$, least squares estimator of β given by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \text{ provided } X^T X \text{ is invertible.}$$

Multicollinearity

Problems arise if predictors are correlated among themselves.

1. If predictors are correlated, X may not be of full rank and thus $X^t X$ may not be invertible.

2. Inflation of variance of $\hat{\beta}$.

3. The least squares estimator $\hat{\beta}$ may be far from true value of β.

Multicollinearity

For simplicity, take a simple linear regression model with two predictors:

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

With x_1, x_2, y standardized.

Then, $X^T X = n \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$ with r the correlation coefficient between predictors x_1 and x_2.

Multicollinearity

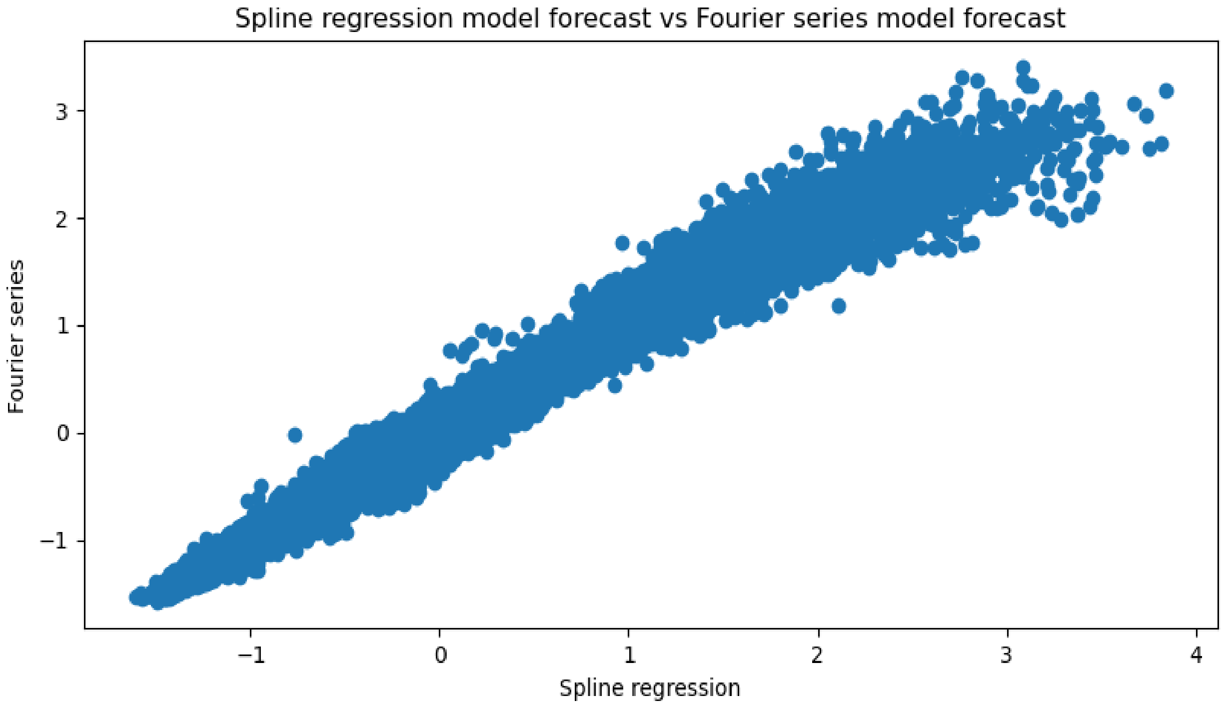
1. We have $|X^T X| = \begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix} = 1 - r^2$. For $r \to 1, |X^t X| \to 0$

2. We can show that $var(\hat{\beta}_i) = \frac{\sigma^2}{n(1-r^2)}$. For $r \to 1, \;\; var(\hat{\beta}_i) \to \infty$

3. We can show that $E\left(\left\|\hat{\beta} - \beta\right\|^2\right) = \frac{2\sigma^2}{1-r^2}$

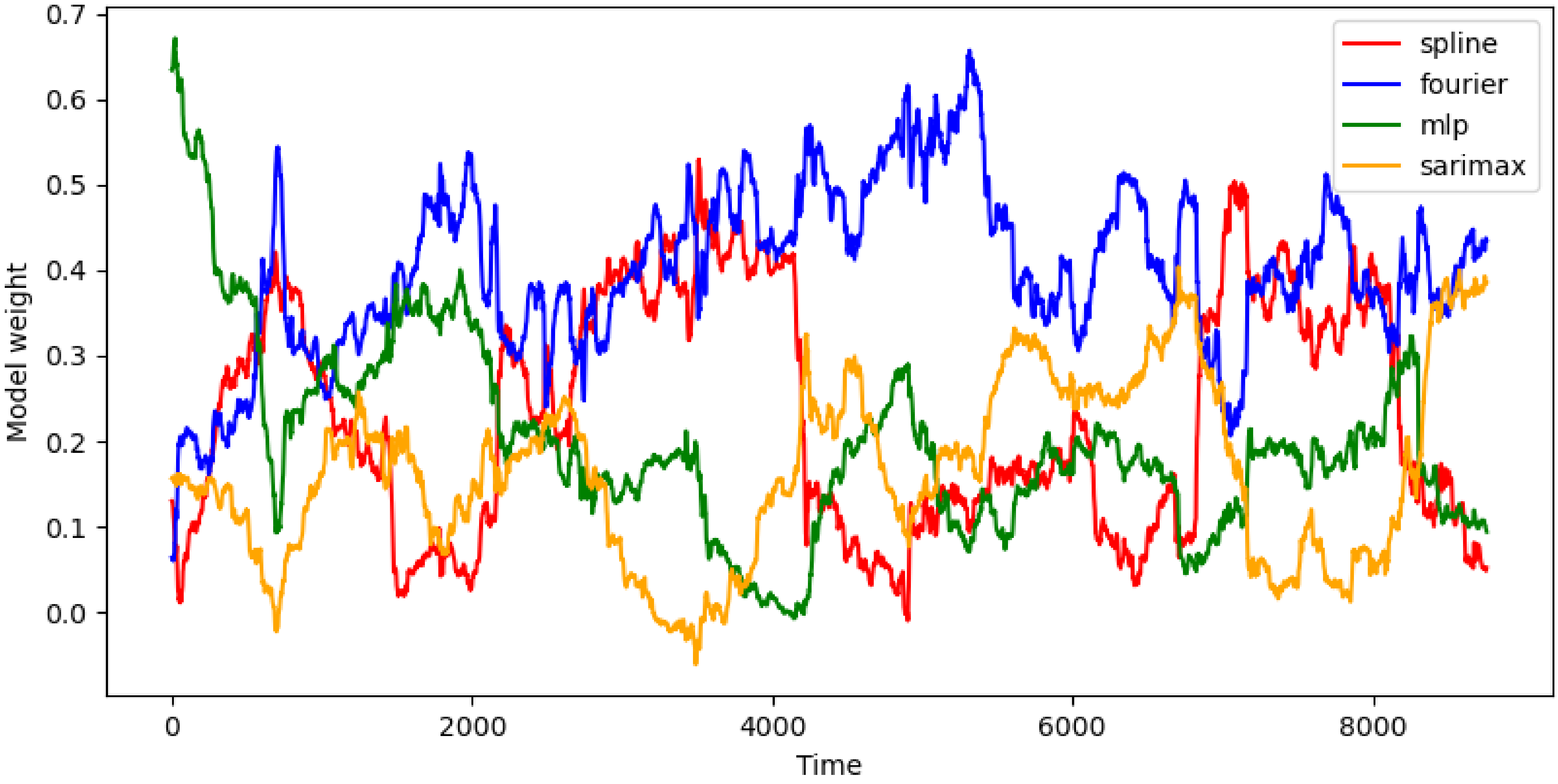
   As $r \to 1, \hat{\beta}$ is on average far from β.

Multicollinearity



- For a time window of 30 days, correlation coefficient between any two models varies between 0.90 and 0.99.

Model weights from OLS ensemble



Regularization

Consider the regularized least squares problem:

$$\min_{\beta} \frac{1}{2} \|Y - X\beta\|_2^2 + \frac{\lambda}{2} \|\beta\|_2^2 = \min_{\beta} \left\| \begin{pmatrix} X \\ \sqrt{\lambda} I \end{pmatrix} \beta - \begin{pmatrix} Y \\ 0 \end{pmatrix} \right\|_2^2$$

Solve like OLS but for $\lambda > 0, \begin{pmatrix} X \\ \sqrt{\lambda} I \end{pmatrix}$ has full rank,

Thus, from normal equations, $\hat{\beta}_r = (X^T X + \lambda I)^{-1} X^T Y$

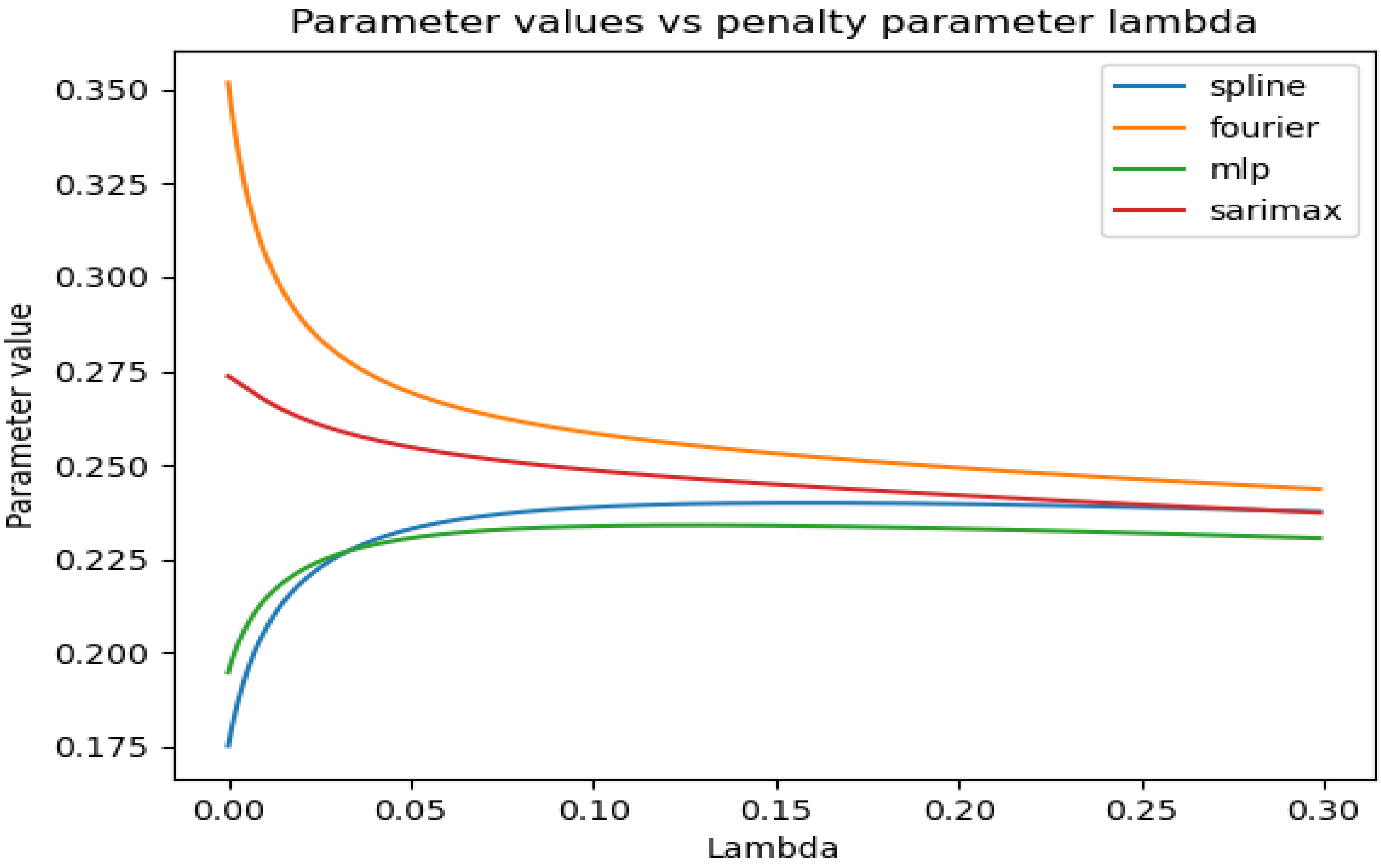
Regularization

Thus, from normal equations, $\hat{\beta}_r = (X^T X + \lambda I)^{-1} X^T Y$

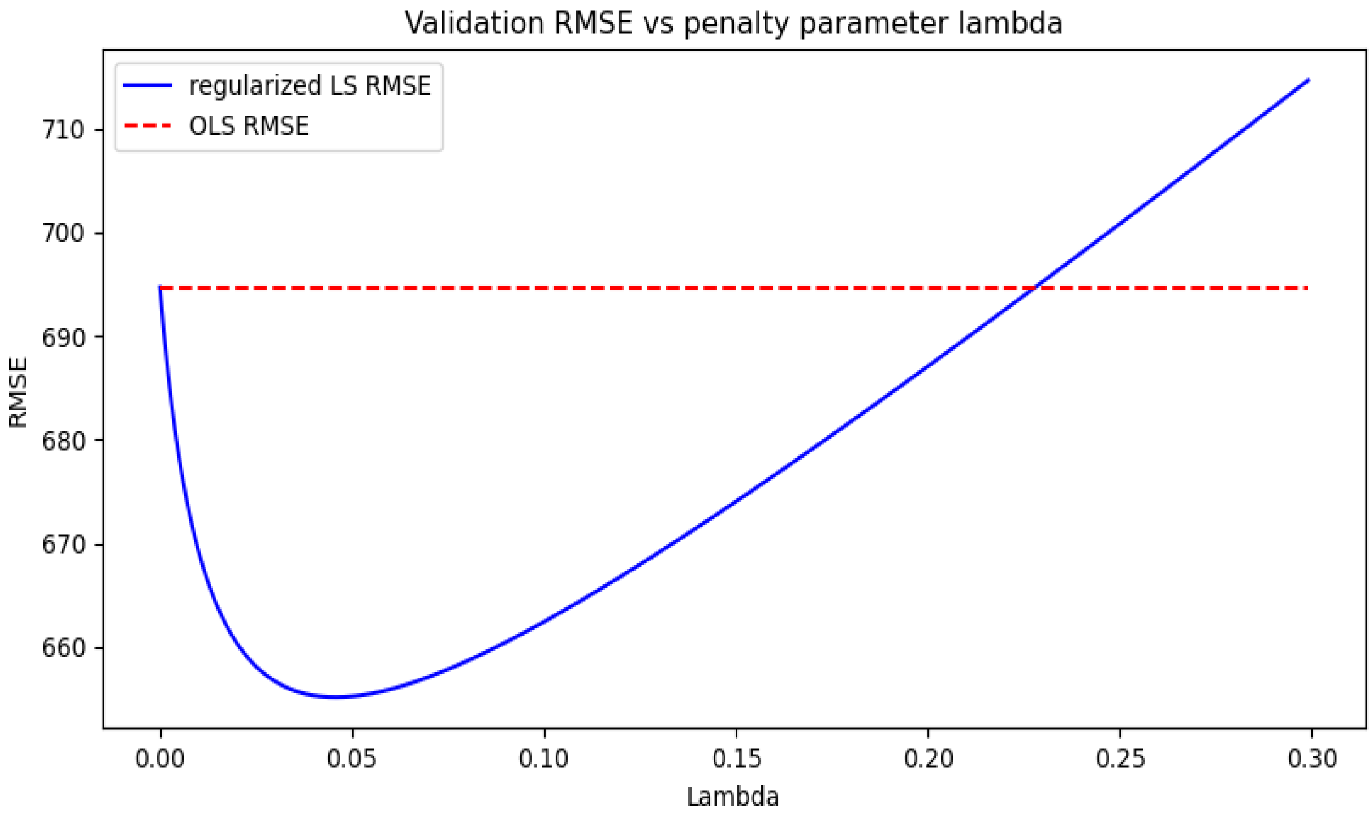
1. Numerically easier to solve for $\hat{\beta}_r$ since $(X^T X + \lambda I)$ is always invertible.

2. We are essentially minimizing $RSS + \lambda ||\beta||_2^2$, by choosing an appropriate λ, we get a tradeoff between magnitude of the parameters and fit.

Regularization



Regularization – Determining λ



- We do not know a priori the optimal λ.
- Take mean optimal lambda from training set.

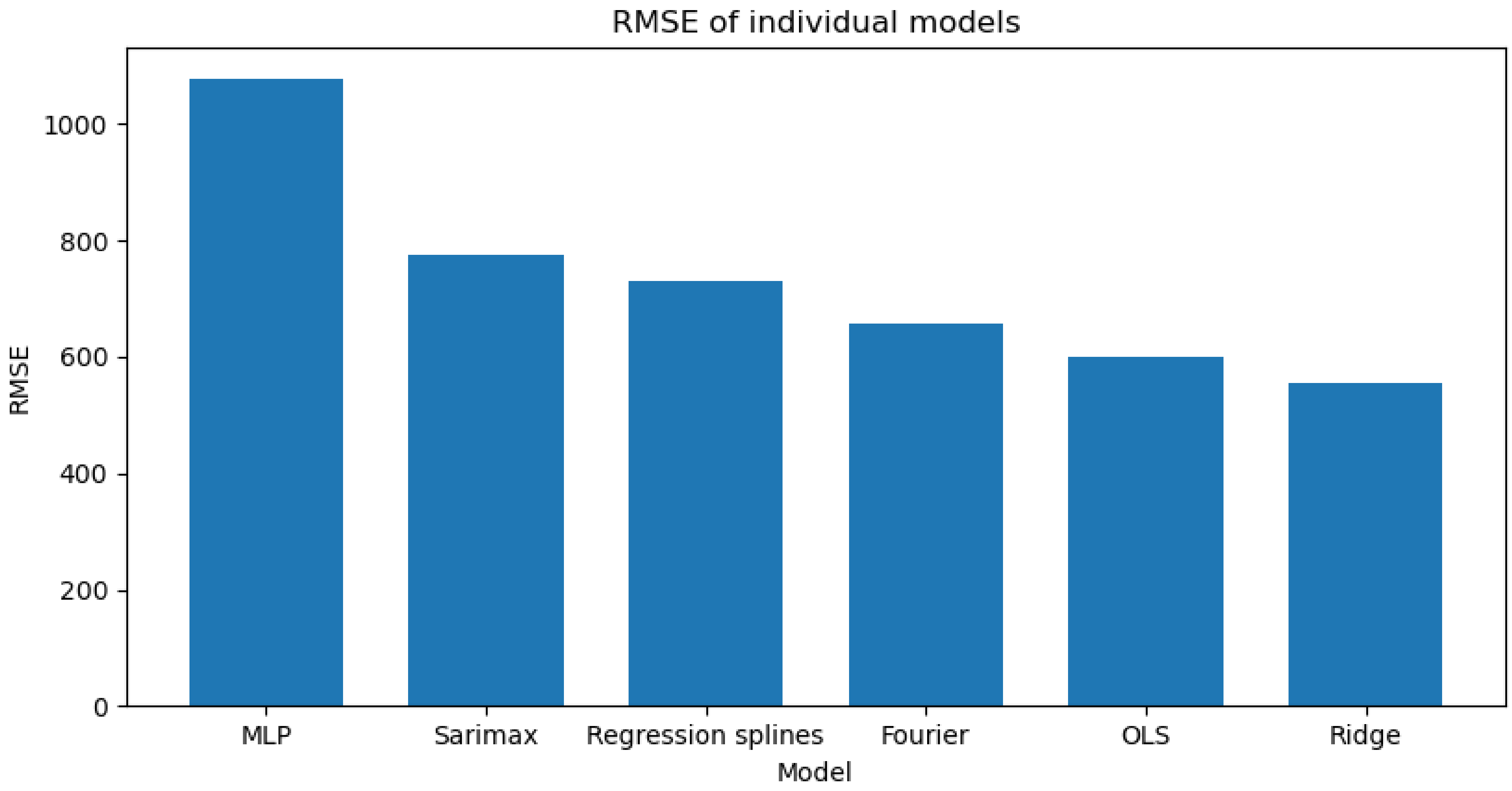
Results – Regularized ensemble method

- Best model – Fourier series
  - RMSE: 656 MW
  - MAPE: 2.24 %

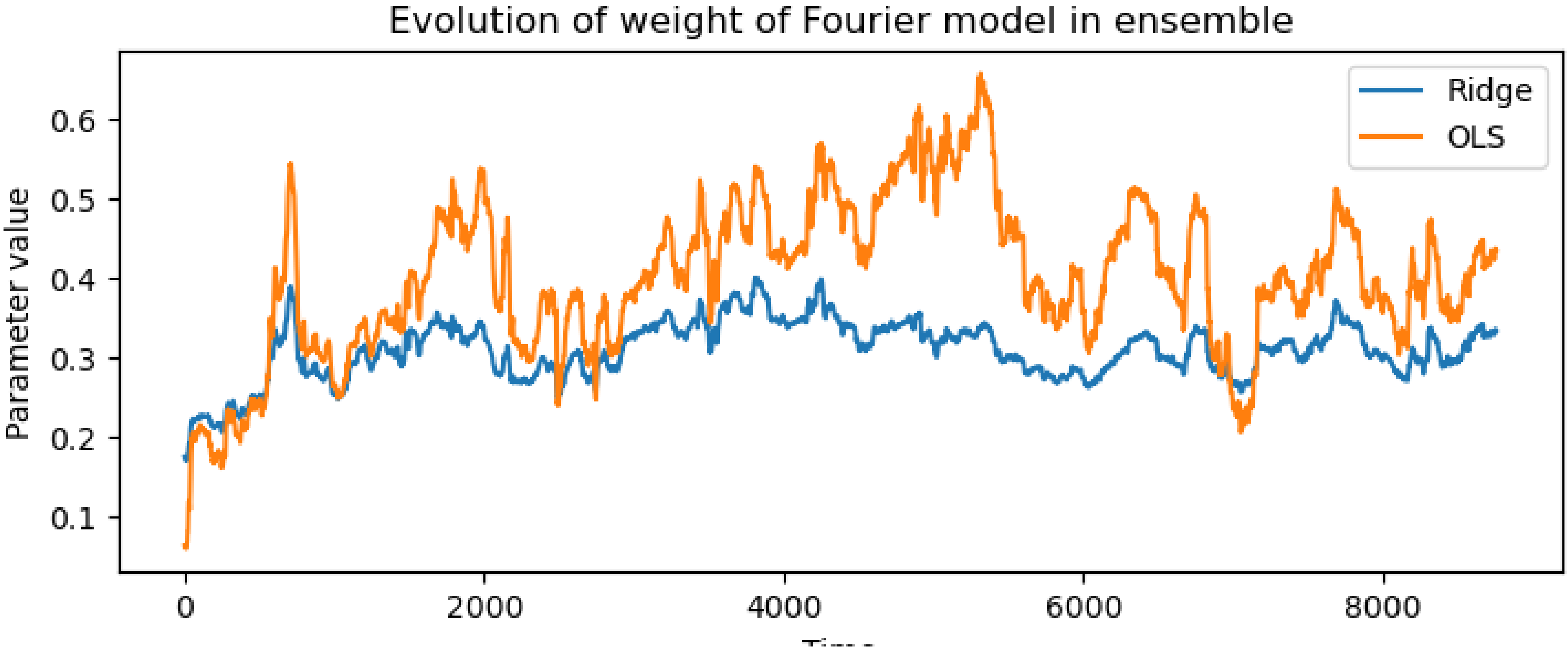
- OLS ensemble method
  - RMSE: 601 MW
  - MAPE: 1.95 %

- Ridge ensemble method
  - RMSE: 553 MW
  - MAPE: 1.82 %

Results

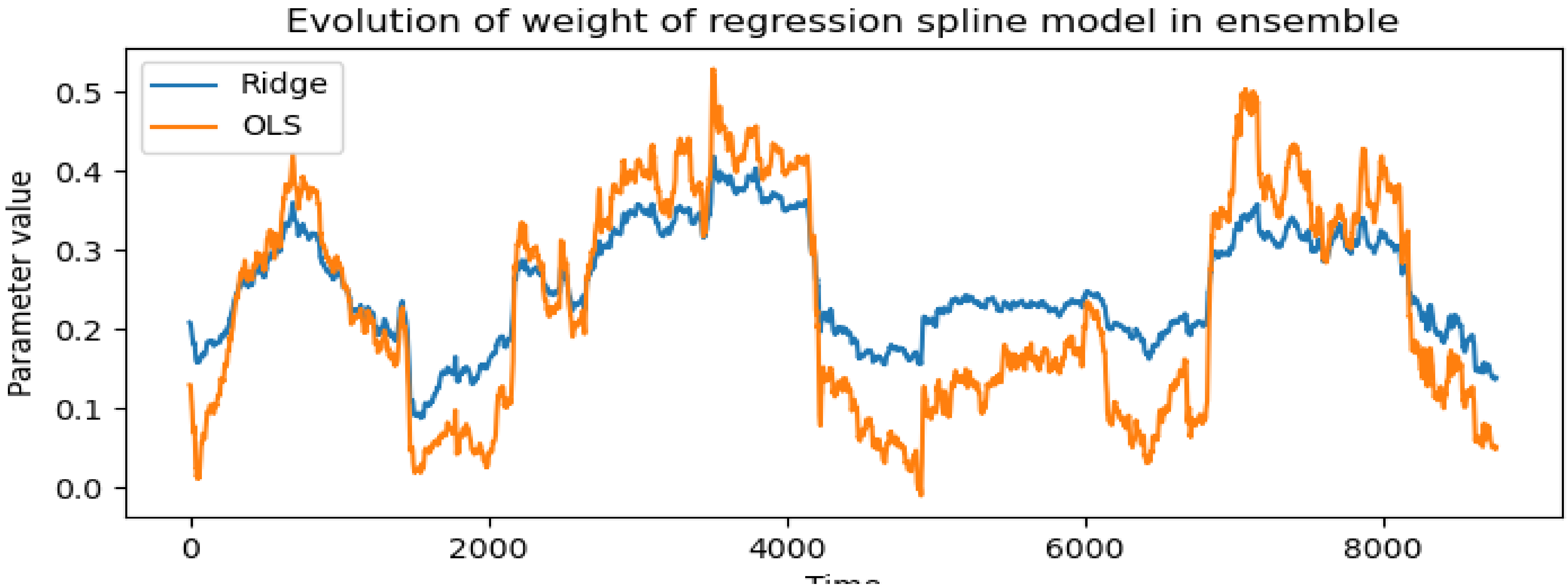

RMSE of individual models
RMSE
0
200
400
600
800
1000
MLP
Sarimax
Regression splines
Fourier
OLS
Ridge
Model

Results



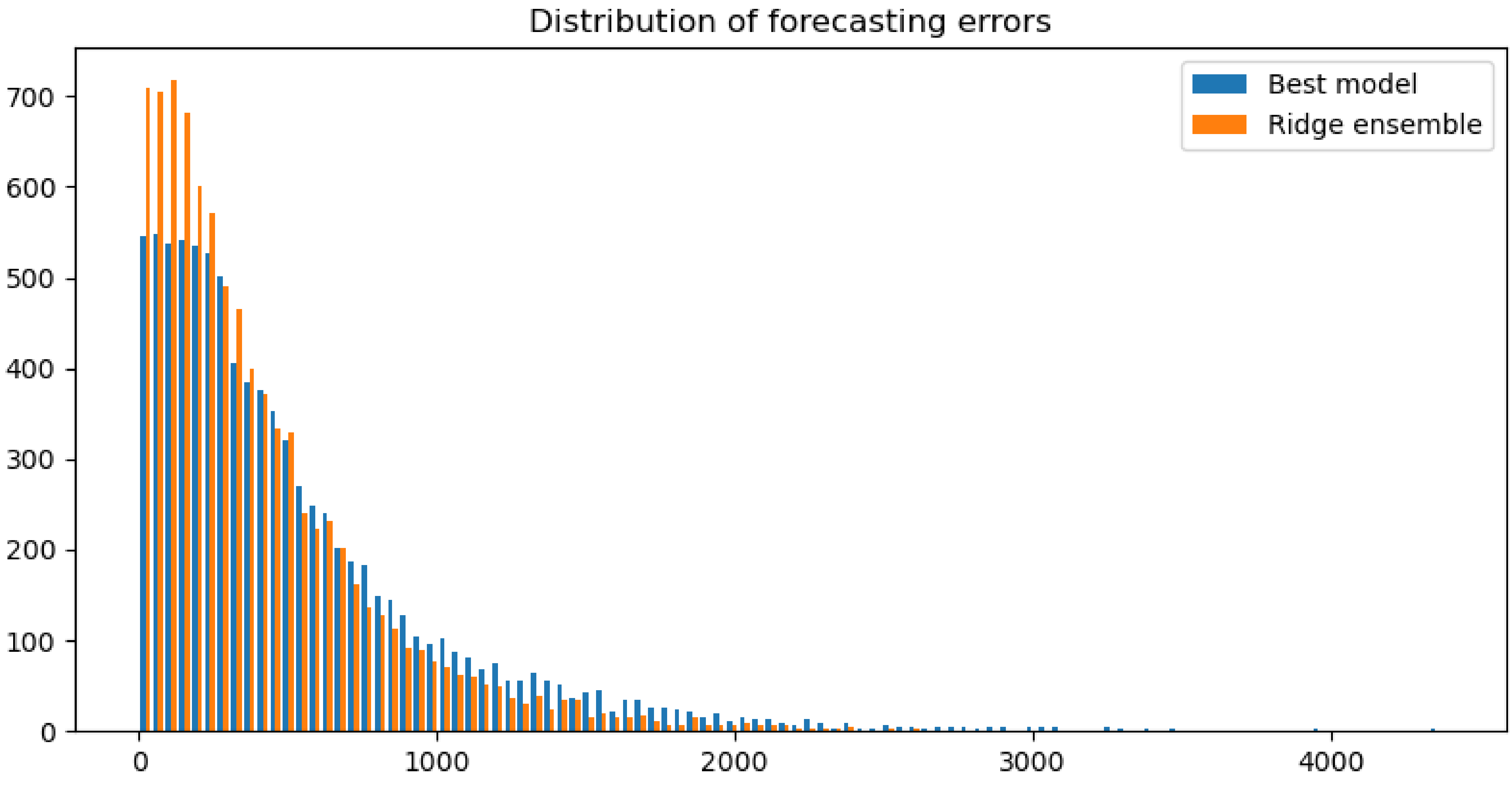
Evolution of weight of Fourier model in ensemble
Ridge
OLS
Parameter value
0.1
0.2
0.3
0.4
0.5
0.6
0
2000
4000
6000
8000

Results



Evolution of weight of regression spline model in ensemble
Ridge
OLS
Parameter value
0.0
0.1
0.2
0.3
0.4
0.5
0
2000
4000
6000
8000

Errors


Distribution of forecasting errors
Best model
Ridge ensemble
0
100
200
300
400
500
600
700
0
1000
2000
3000
4000