Forecast reconciliation with subset selection

Table of contents

1 Group best-subset selection

1.1 MinT reconciliation

The unique solution of MinT is $G = \left(S'W_h^{-1}S\right)^{-1} S'W_h^{-1}$, which has a similar representation to a GLS estimator of a least square problem.

Consider a hierarchy consisting of n time series in total and n_b time series in the bottom level. Let $y_t \in \mathbb{R}^n$ denote a vector of observations at time t of all time series in the hierarchy, $b_t \in \mathbb{R}^{n_b}$ $(n_b < n)$ denote a vector of observations at time t of only the most disaggregated bottom-level series.

Therefore, the trace minimization problem can be reformulated in terms of a Quadratic Programming (QP) problem as follows:

$$
\begin{aligned}
\min_{G\hat{y}_h} \quad & \left(\hat{y}_h - SG\hat{y}_h\right)' W_h^{-1} \left(\hat{y}_h - SG\hat{y}_h\right) \\
\text{s.t.} \quad & GS = I_{n_b}.
\end{aligned} \tag{1}
$$

Note that the variable of interest is $G\hat{y}_h$ rather than G. So we can get a unique solution of reconciled forecasts at the bottom level, while infinitely many least-squares solutions of G as the columns of $\hat{y}_h' \otimes S$ are not linearly independent.

1.2 Best-subset selection

To eliminate the negative effect of some underperforming base forecasts on the performance of the reconciled forecasts, we want to **zero out some columns of** G. Thus, the corresponding base forecasts in $\hat{y}_h$ are not used to form the reconciled bottom-level forecasts and, moreover, are not used for all reconciled forecasts.

One way to achieve this goal is by considering an ℓ_0-norm regularization. **Best-subset selection** generally performs well in high signal-to-noise (SNR) ratio regimes, while lasso performs better in low SNR regimes. We also include an additional ℓ_2-norm regularization (in addition to the ℓ_0 penalty), which is motivated by some related works (Hastie, Tibshirani, and Tibshirani 2020; Mazumder, Radchenko, and Dedieu 2023), which suggest that when the SNR is low, additional ridge regularization can improve the prediction performance of best-subset selection.

1.3 Group best-subset selection with ridge regularization

The vectorization is frequently used together with the Kronecker product to express matrix multiplication as a linear transformation on matrices $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$. Therefore, the QP minimization problem can be reduced to a regression problem as follows:

$$\begin{aligned} \min_{G} \quad & \frac{1}{2}\left(\hat{y}_h - \left(\hat{y}_h' \otimes S\right)\operatorname{vec}(G)\right)' W_h^{-1}\left(\hat{y}_h - \left(\hat{y}_h' \otimes S\right)\operatorname{vec}(G)\right) \\ \text{s.t.} \quad & GS = I_{n_b}. \end{aligned} \tag{2}$$

Here, we consider the following $\ell_0\ell_2$**-regularized regression problem** of the following form to achieve selection in hierarchical forecasting:

$$\begin{aligned} \min_{G} \quad & \frac{1}{2}\left(\hat{y} - \left(\hat{y}' \otimes S\right)\operatorname{vec}(G)\right)' W^{-1}\left(\hat{y} - \left(\hat{y}' \otimes S\right)\operatorname{vec}(G)\right) \\ & \quad + \lambda_0 \sum_{j=1}^{n} \left\|G_{\cdot j}\right\|_0 + \lambda_2 \left\|\operatorname{vec}(G)\right\|_2^2 \\ \text{s.t.} \quad & GS = I_{n_b}, \end{aligned} \tag{3}$$

where $\lambda_0 > 0$ controls the number of non-zero columns of G, and $\lambda_2 \geqslant 0$ controls the strength of the ridge regularization, $\sum_{j=1}^{n} \left\|G_{\cdot j}\right\|_0$ is the number of non-zero columns of G. In a hierarchy setting, the target variable, $\operatorname{vec}(G)$, in the minimization problem has a natural group structure, i.e., each column of G is a group. Thus, the $n \times n_b$ predictors in the regularized regression problem are divided into n pre-specified, non-overlapping groups, with each group consisting of n_b predictors. Therefore, the target problem is essentially **a group best-subset selection with ridge regularization**.

1.4 Mixed integer program

We propose MIP formulations to solve Equation 3. We first present a **Big-M based MIP formulation** for problem Equation 3:

$$
\begin{aligned}
&\min_{G,z,\check{e},g^+} \frac{1}{2}\check{e}' W_h^{-1}\check{e} + \lambda_0 \sum_{j=1}^{n} z_j + \lambda_2 g^{+\prime} g^+ \\
&\text{s.t.} \quad \hat{y}_h - \left(\hat{y}_h' \otimes S\right)\operatorname{vec}(G) = \check{e} \quad \cdots (C1) \\
&\qquad\quad GS = I_{n_b} \Leftrightarrow \left(S' \otimes I_{n_b}\right)\operatorname{vec}(G) = \operatorname{vec}\left(I_{n_b}\right) \quad \cdots (C2) \\
&\qquad\quad \sum_{i=1}^{n_b} g^+_{i+(j-1)n_b} \leqslant \mathcal{M} z_j, \quad j \in [n] \quad \cdots (C3) \\
&\qquad\quad g^+ \geqslant \operatorname{vec}(G) \quad \cdots (C4) \\
&\qquad\quad g^+ \geqslant -\operatorname{vec}(G) \quad \cdots (C5) \\
&\qquad\quad z_j \in \{0,1\}, \quad j \in [n] \quad \cdots (C6)
\end{aligned} \tag{4}
$$

where, $\mathcal{M}$ is a priori specified constant (leading to the name "Big-M") such that some optimal solution, say g^{+*}, to Equation 4 satisfies $\max_{j\in[n]} \sum_{i=1}^{n_b} g^+_{i+(j-1)n_b} \leqslant \mathcal{M}$, the binary variable z_j controls whether all the regression coefficients in group j are zero or not: $z_j = 0$ implies that $G_{\cdot j} = \mathbf{0}$, and $z_j = 1$ implies that $\sum_{i=1}^{n_b} g^+_{i+(j-1)n_b} \leqslant \mathcal{M}$. Such Big-M formulations are commonly used in mixed integer programming to model relations between discrete and continuous variables, and have been recently used in ℓ_0-regularized regression.

This is a **Mixed Integer Quadratic Program (MIQP)** and then get solved using some efficient commercial solvers such as Gurobi, CPLEX, and MOSEK. Note that the best subset selection is an **NP-hard problem**, which is computationally intensive.

1.5 Hyperparameter

- $\lambda_0 = \{0, 10^{k-3}, 10^{k-2}, 10^{k-1}, 10^{k}, 10^{k+1}\}$, where k is the number of digits before the decimal point for $\frac{1}{2n_b}\left(\hat{y}_h - \tilde{y}_h^{\text{MinT}}\right)' W_h^{-1}\left(\hat{y}_h - \tilde{y}_h^{\text{MinT}}\right)$. (Reason)
- $\lambda_2 = \{0, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$

To avoid cross-validation, we select the best combination of λ_0 and λ_2 by minimizing the sum of squared reconciled forecast errors in the training set, even though fitted values are often not true one-step ahead forecasts.

1.6 Simulation results

1.6.1 Data simulation

Structure:

- Top: Total

- Middle: A, B
- Bottom: AA, AB, BA, BB

Data generation:

The bottom-level series were generated using the basic structural time series model

$$b_t = \mu_t + \gamma_t + \eta_t$$

where μ_t, γ_t, and η_t are the trend, seasonal, and error components, respectively,

$$\begin{aligned}
\mu_t &= \mu_{t-1} + v_t + \varrho_t, & \varrho_t &\sim \mathcal{N}\left(\mathbf{0}, \sigma_\varrho^2 I_4\right), \\
v_t &= v_{t-1} + \zeta_t, & \zeta_t &\sim \mathcal{N}\left(\mathbf{0}, \sigma_\zeta^2 I_4\right), \\
\gamma_t &= -\sum_{i=1}^{s-1} \gamma_{t-i} + \omega_t, & \omega_t &\sim \mathcal{N}\left(\mathbf{0}, \sigma_\omega^2 \mathbf{I}_4\right),
\end{aligned}$$

and ϱ_t, ζ_t, and ω_t are errors independent of each other and over time.

Other details:

- $\sigma_\varrho^2 = 2, \sigma_\zeta^2 = 0.007$, and $\sigma_\omega^2 = 7$.
- $s = 4$ for quarterly data, $n = 180$, $h = 16$.
- The initial values for $\mu_0, v_0, \gamma_0, \gamma_1, \gamma_2$ were generated independently from a multivariate normal distribution with mean zero and covariance matrix, $\Sigma_0 = I_4$.
- Each component of η_t was generated from an ARIMA$(p, 0, q)$ process with p and q taking values of 0 and 1 with equal probability.
- The bottom-level series were then appropriately summed to obtain the data for higher levels.
- This process was repeated 500 times.

1.6.2 Results

Scenario 0: ETS

- ETS models are used to generate base forecasts. Table 1, Table 2, and Figure 1.

Table 1: Out-of-sample average RMSE results in Scenario 0.

Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	9.61	10.73	12.59	15.58	6.33	7.26	8.61	10.83	4.20	4.92	5.93	7.52	5.58	6.41	7.65	9.62
BU	9.51	10.78	12.67	15.68	6.32	7.25	8.62	10.83	4.20	4.92	5.93	7.52	5.56	6.42	7.66	9.63
OLS	9.54	10.71	12.59	15.59	6.33	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.57	6.40	7.63	9.60
OLS-subset	9.54	10.73	12.61	15.61	6.32	7.24	8.60	10.82	4.20	4.91	5.93	7.52	5.57	6.41	7.65	9.62
WLSs	9.52	10.71	12.60	15.60	6.32	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
WLSs-subset	9.54	10.73	12.61	15.60	6.32	7.24	8.60	10.81	4.20	4.91	5.93	7.52	5.57	6.41	7.65	9.61
WLSv	9.52	10.72	12.60	15.61	6.32	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
WLSv-subset	9.53	10.72	12.61	15.61	6.31	7.24	8.60	10.81	4.20	4.91	5.93	7.52	5.57	6.41	7.64	9.61
MinT	9.54	10.74	12.62	15.62	6.31	7.25	8.61	10.82	4.22	4.92	5.93	7.52	5.58	6.42	7.65	9.62
MinT-subset	9.54	10.74	12.62	15.62	6.31	7.25	8.61	10.82	4.22	4.92	5.93	7.52	5.58	6.42	7.65	9.62
MinTs	9.52	10.72	12.60	15.60	6.31	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
MinTs-subset	9.52	10.72	12.60	15.60	6.31	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61

Table 2: Out-of-sample average MASE results in Scenario 0.

Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	0.91	0.87	0.99	1.20	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	0.90	0.88	1.02	1.27
BU	0.90	0.88	1.00	1.21	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	0.89	0.88	1.03	1.27
OLS	0.91	0.87	0.99	1.20	0.90	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.90	0.88	1.02	1.27
OLS-subset	0.91	0.87	0.99	1.21	0.90	0.87	1.01	1.25	0.89	0.88	1.04	1.30	0.90	0.88	1.02	1.27
WLSs	0.91	0.87	0.99	1.21	0.90	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
WLSs-subset	0.91	0.87	0.99	1.21	0.90	0.87	1.01	1.25	0.89	0.88	1.04	1.30	0.89	0.88	1.02	1.27
WLSv	0.91	0.87	0.99	1.21	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
WLSv-subset	0.91	0.87	0.99	1.21	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.30	0.89	0.88	1.02	1.27
MinTs	0.91	0.87	0.99	1.20	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
MinTs-subset	0.91	0.87	0.99	1.20	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27

Figure 1: Frequency of the base forecasts being removed from reconciliation in Scenario 0.

Scenario 1: D-AA

- Base forecasts (and also fitted values) of **series AA** multiplied by 1.5 to achieve deterioration. Table 3, Table 4, and Figure 2.

Table 3: Out-of-sample average RMSE results in Scenario 1.

	Top				Middle				Bottom				Average			
Method	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	9.61	10.73	12.59	15.58	6.33	7.26	8.61	10.83	6.38	7.47	8.34	9.75	6.83	7.88	9.02	10.89
BU	15.16	18.08	19.35	21.64	10.02	11.74	12.75	14.55	6.38	7.47	8.34	9.75	8.68	10.21	11.17	12.82
OLS	9.66	10.96	12.82	15.80	6.78	7.72	9.00	11.16	5.90	6.83	7.66	9.04	6.69	7.68	8.78	10.61
OLS-subset	9.52	10.73	12.60	15.61	6.35	7.29	8.63	10.83	4.30	5.05	6.05	7.62	5.63	6.50	7.72	9.68
WLSs	10.31	11.86	13.62	16.50	7.32	8.42	9.62	11.70	5.94	6.89	7.72	9.12	6.96	8.04	9.11	10.91
WLSs-subset	10.06	11.26	13.10	16.05	7.09	7.92	9.19	11.34	5.86	6.57	7.43	8.85	6.81	7.62	8.74	10.59
WLSv	9.70	11.04	12.89	15.87	6.62	7.57	8.88	11.06	4.74	5.50	6.44	7.97	5.98	6.88	8.06	9.98
WLSv-subset	9.50	10.75	12.65	15.64	6.33	7.27	8.63	10.85	4.24	4.99	6.00	7.58	5.59	6.47	7.70	9.66
MinT	9.57	10.81	12.70	15.67	6.38	7.31	8.66	10.86	4.28	4.98	5.98	7.56	5.64	6.48	7.71	9.66
MinT-subset	9.57	10.81	12.70	15.67	6.38	7.31	8.66	10.86	4.28	4.98	5.98	7.56	5.64	6.48	7.71	9.66
MinTs	9.52	10.79	12.69	15.66	6.37	7.30	8.65	10.85	4.28	4.97	5.98	7.56	5.63	6.47	7.70	9.66
MinTs-subset	9.52	10.79	12.68	15.66	6.37	7.30	8.65	10.85	4.28	4.97	5.98	7.56	5.63	6.47	7.70	9.66

Table 4: Out-of-sample average MASE results in Scenario 1.

Method	Top h=1	Top 1-4	Top 1-8	Top 1-16	Middle h=1	Middle 1-4	Middle 1-8	Middle 1-16	Bottom h=1	Bottom 1-4	Bottom 1-8	Bottom 1-16	Average h=1	Average 1-4	Average 1-8	Average 1-16
Base	0.91	0.87	0.99	1.20	0.90	0.88	1.01	1.25	1.35	1.34	1.48	1.71	1.16	1.14	1.27	1.50
BU	1.44	1.48	1.56	1.72	1.42	1.42	1.52	1.71	1.35	1.34	1.48	1.71	1.38	1.38	1.50	1.71
OLS	0.92	0.89	1.01	1.22	0.96	0.93	1.06	1.29	1.25	1.23	1.36	1.58	1.12	1.10	1.22	1.45
OLS-subset	0.91	0.87	0.99	1.20	0.90	0.88	1.01	1.25	0.91	0.91	1.06	1.31	0.91	0.89	1.04	1.28
WLSs	0.98	0.96	1.08	1.28	1.04	1.02	1.14	1.36	1.26	1.24	1.37	1.60	1.15	1.14	1.26	1.48
WLSs-subset	0.96	0.91	1.03	1.25	1.00	0.96	1.08	1.31	1.24	1.18	1.32	1.55	1.13	1.08	1.21	1.44
WLSv	0.92	0.90	1.01	1.23	0.94	0.92	1.05	1.28	1.00	0.99	1.13	1.38	0.97	0.95	1.09	1.33
WLSv-subset	0.90	0.87	0.99	1.21	0.90	0.88	1.01	1.25	0.90	0.90	1.05	1.31	0.90	0.89	1.03	1.28
MinTs	0.91	0.87	1.00	1.21	0.90	0.88	1.02	1.25	0.91	0.89	1.05	1.30	0.91	0.89	1.03	1.28
MinTs-subset	0.91	0.87	1.00	1.21	0.90	0.88	1.02	1.25	0.91	0.89	1.05	1.30	0.91	0.89	1.03	1.28

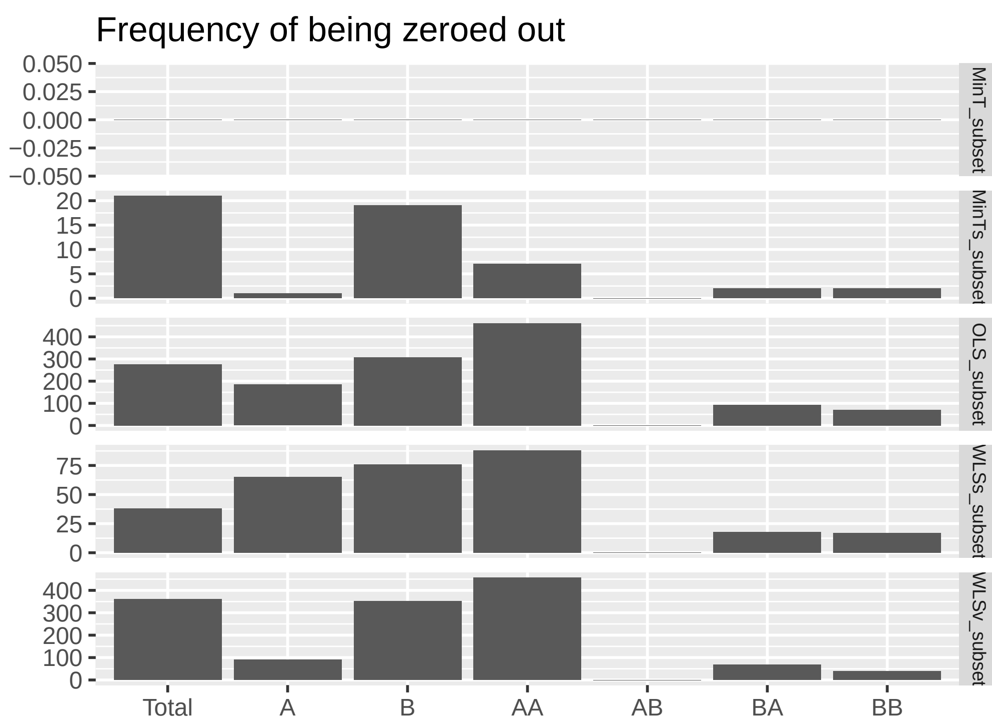

Figure 2: Frequency of the base forecasts being removed from reconciliation in Scenario 1.

Scenario 2: D-A

- Base forecasts (and also fitted values) of **series A** multiplied by 1.5 to achieve deterioration. Table 5, Table 6, and Figure 3.

Table 5: Out-of-sample average RMSE results in Scenario 2.

Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	9.61	10.73	12.59	15.58	12.07	14.38	15.29	16.97	4.20	4.92	5.93	7.52	7.22	8.45	9.56	11.37
BU	9.51	10.78	12.67	15.68	6.32	7.25	8.62	10.83	4.20	4.92	5.93	7.52	5.56	6.42	7.66	9.63
OLS	10.42	12.22	13.91	16.76	8.67	10.16	11.20	13.04	5.16	6.09	6.94	8.37	6.91	8.13	9.15	10.90
OLS-subset	9.48	10.77	12.66	15.66	6.32	7.27	8.63	10.84	4.21	4.93	5.95	7.54	5.56	6.43	7.67	9.64
WLSs	10.77	12.73	14.36	17.17	7.92	9.33	10.45	12.40	4.85	5.75	6.64	8.12	6.57	7.77	8.83	10.64
WLSs-subset	9.54	10.77	12.66	15.67	6.33	7.27	8.64	10.84	4.21	4.93	5.95	7.54	5.58	6.43	7.68	9.64
WLSv	9.53	10.97	12.82	15.83	6.48	7.50	8.82	11.00	4.27	5.01	6.00	7.58	5.65	6.57	7.78	9.74
WLSv-subset	9.55	10.81	12.70	15.69	6.35	7.31	8.67	10.86	4.21	4.94	5.95	7.54	5.58	6.45	7.69	9.65
MinT	9.62	10.78	12.67	15.65	6.33	7.28	8.65	10.84	4.24	4.94	5.95	7.53	5.61	6.44	7.68	9.64
MinT-subset	9.62	10.78	12.67	15.65	6.33	7.28	8.65	10.84	4.24	4.94	5.95	7.53	5.61	6.44	7.68	9.64
MinTs	9.57	10.76	12.65	15.64	6.32	7.26	8.63	10.83	4.23	4.93	5.94	7.52	5.59	6.43	7.67	9.63
MinTs-subset	9.57	10.76	12.65	15.64	6.32	7.26	8.63	10.83	4.23	4.93	5.94	7.52	5.59	6.43	7.66	9.63

Table 6: Out-of-sample average MASE results in Scenario 2.

Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	0.91	0.87	0.99	1.20	1.71	1.73	1.82	2.01	0.89	0.88	1.04	1.30	1.13	1.12	1.26	1.49
BU	0.90	0.88	1.00	1.21	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	0.89	0.88	1.03	1.27
OLS	0.99	0.99	1.10	1.31	1.23	1.23	1.33	1.53	1.09	1.10	1.23	1.46	1.12	1.12	1.24	1.46
OLS-subset	0.90	0.87	0.99	1.21	0.90	0.88	1.01	1.25	0.89	0.89	1.04	1.30	0.89	0.88	1.03	1.27
WLSs	1.02	1.03	1.14	1.34	1.12	1.13	1.24	1.44	1.03	1.04	1.17	1.41	1.05	1.06	1.19	1.41
WLSs-subset	0.91	0.87	0.99	1.21	0.90	0.88	1.01	1.25	0.89	0.89	1.04	1.30	0.90	0.88	1.03	1.27
WLSv	0.91	0.89	1.01	1.22	0.92	0.91	1.04	1.27	0.91	0.90	1.05	1.31	0.91	0.90	1.04	1.28
WLSv-subset	0.91	0.88	1.00	1.21	0.90	0.88	1.02	1.25	0.89	0.89	1.04	1.30	0.90	0.89	1.03	1.27
MinTs	0.91	0.87	0.99	1.21	0.90	0.88	1.01	1.25	0.90	0.89	1.04	1.30	0.90	0.88	1.03	1.27
MinTs-subset	0.91	0.87	0.99	1.21	0.90	0.88	1.01	1.25	0.90	0.89	1.04	1.30	0.90	0.88	1.03	1.27

Figure 3: Frequency of the base forecasts being removed from reconciliation in Scenario 2.

Scenario 3: D-Total

- Base forecasts (and also fitted values) of **series Total** multiplied by 1.5 to achieve deterioration. Table 7, Table 8 and Figure 4.

Table 7: Out-of-sample average RMSE results in Scenario 3.

	Top				Middle				Bottom				Average			
Method	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	25.01	30.26	30.88	32.34	6.33	7.26	8.61	10.83	4.20	4.92	5.93	7.52	7.78	9.20	10.26	12.01
BU	9.51	10.78	12.67	15.68	6.32	7.25	8.62	10.83	4.20	4.92	5.93	7.52	5.56	6.42	7.66	9.63
OLS	16.30	19.50	20.55	22.59	9.20	10.93	11.85	13.55	5.36	6.39	7.19	8.55	8.02	9.56	10.43	11.99
OLS-subset	9.48	10.76	12.67	15.69	6.31	7.26	8.63	10.85	4.20	4.92	5.93	7.53	5.55	6.42	7.67	9.64
WLSs	12.27	14.39	15.84	18.35	7.45	8.71	9.86	11.83	4.60	5.47	6.39	7.89	6.51	7.67	8.73	10.51
WLSs-subset	9.48	10.76	12.66	15.68	6.30	7.26	8.62	10.84	4.19	4.92	5.94	7.53	5.55	6.42	7.66	9.64
WLSv	9.72	11.08	12.95	15.91	6.40	7.38	8.72	10.91	4.23	4.96	5.97	7.55	5.63	6.53	7.75	9.71
WLSv-subset	9.51	10.83	12.71	15.72	6.33	7.29	8.64	10.85	4.20	4.93	5.94	7.53	5.57	6.45	7.68	9.65
MinT	9.48	10.81	12.68	15.66	6.32	7.30	8.65	10.85	4.23	4.94	5.95	7.53	5.58	6.45	7.68	9.64
MinT-subset	9.48	10.81	12.68	15.66	6.32	7.30	8.65	10.85	4.23	4.94	5.95	7.53	5.58	6.45	7.68	9.64
MinTs	9.46	10.78	12.66	15.65	6.32	7.28	8.64	10.84	4.21	4.93	5.94	7.52	5.56	6.44	7.67	9.63
MinTs-subset	9.46	10.78	12.66	15.65	6.32	7.28	8.64	10.83	4.21	4.93	5.94	7.52	5.56	6.44	7.67	9.63

Table 8: Out-of-sample average MASE results in Scenario 3.

	Top				Middle				Bottom				Average			
Method	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	2.38	2.45	2.49	2.60	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	1.11	1.11	1.24	1.47
BU	0.90	0.88	1.00	1.21	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	0.89	0.88	1.03	1.27
OLS	1.55	1.58	1.65	1.80	1.30	1.32	1.41	1.59	1.14	1.15	1.27	1.49	1.24	1.26	1.37	1.57
OLS-subset	0.90	0.87	1.00	1.21	0.89	0.88	1.01	1.25	0.89	0.89	1.04	1.30	0.89	0.88	1.03	1.27
WLSs	1.17	1.17	1.26	1.44	1.06	1.05	1.17	1.38	0.98	0.99	1.13	1.37	1.03	1.03	1.16	1.38
WLSs-subset	0.90	0.87	0.99	1.21	0.89	0.88	1.01	1.25	0.89	0.89	1.04	1.30	0.89	0.88	1.03	1.27
WLSv	0.93	0.90	1.02	1.23	0.91	0.89	1.02	1.26	0.90	0.89	1.05	1.30	0.90	0.89	1.04	1.28
WLSv-subset	0.90	0.88	1.00	1.22	0.90	0.88	1.02	1.25	0.89	0.89	1.04	1.30	0.90	0.88	1.03	1.27
MinTs	0.90	0.88	0.99	1.21	0.90	0.88	1.01	1.25	0.89	0.89	1.04	1.30	0.90	0.88	1.03	1.27
MinTs-subset	0.90	0.88	0.99	1.21	0.89	0.88	1.01	1.25	0.89	0.89	1.04	1.30	0.89	0.88	1.03	1.27

Figure 4: Frequency of the base forecasts being removed from reconciliation in Scenario 3.

1.7 Tourism data results

1.7.1 Data description

Australian domestic tourism (only considering hierarchical structure)

- Monthly series from 1998 Jan to 2017 Dec.
- 240 months (20 years) for each series.
- Hierarchy: Total/State/Zone/Region, 4 levels, n = 111 series in total.
- Training set: 1998 Jan-2016 Dec.
- Test set: 2017 Jan-2017 Dec.

1.7.2 Results

- OLS_subset: $\lambda_0 = 0, \lambda_2 = 0.1, \sum z = 111$
- WLSs_subset: $\lambda_0 = 10, \lambda_2 = 10, \sum z = 92$
- WLSv_subset: $\lambda_0 = 0, \lambda_2 = 100, \sum z = 111$
- MinTs_subset: $\lambda_0 = 0, \lambda_2 = 0, \sum z = 111$

Table 9: Out-of-sample average RMSE results in Tourism data.

	Top				Sate				Zone				Region				Average			
Method	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12
Base	1158.16	716.62	1279.50	1907.61	452.68	323.31	349.92	424.77	165.52	163.62	160.69	179.71	100.80	89.43	88.25	94.11	148.26	127.87	133.11	152.12
BU	2189.97	1667.97	1962.73	2708.64	431.95	356.46	409.53	508.37	167.29	159.70	161.40	181.55	100.80	89.43	88.25	94.11	156.68	137.58	143.19	165.06
OLS	1103.20	714.05	1286.29	1935.26	438.89	310.65	344.21	418.35	162.12	156.78	151.75	166.20	101.79	89.07	86.56	91.13	146.75	125.14	129.48	146.64
OLS-subset	1103.20	714.05	1286.29	1935.26	438.89	310.65	344.21	418.35	162.12	156.78	151.75	166.20	101.79	89.07	86.56	91.13	146.75	125.14	129.48	146.64
WLSs	1448.95	1111.95	1546.09	2271.59	381.21	307.05	362.27	451.23	155.70	154.71	153.10	170.80	100.60	88.72	86.85	92.07	143.85	127.76	133.48	153.51
WLSs-subset	1448.83	857.56	1360.82	1991.27	381.19	308.78	341.59	415.61	155.69	153.91	150.56	165.36	100.60	88.90	86.90	91.65	143.84	125.51	129.92	147.12
WLSv	1600.65	1262.35	1657.96	2395.97	374.05	313.35	374.44	466.85	157.23	156.62	155.64	173.97	96.61	88.01	86.66	92.16	142.40	129.49	135.74	156.44
WLSv-subset	1329.28	826.08	1343.27	1995.38	412.37	302.00	343.91	421.96	159.42	155.94	151.77	166.75	101.26	88.91	86.48	91.16	146.09	125.30	129.92	147.56
MinTs	1397.23	1101.06	1555.74	2270.36	352.16	300.15	362.01	451.41	145.46	152.82	152.55	170.11	95.41	87.07	85.82	91.15	135.50	125.63	132.71	152.71
MinTs-subset	1397.23	1101.06	1555.74	2270.36	352.16	300.15	362.01	451.41	145.46	152.82	152.55	170.11	95.41	87.07	85.82	91.15	135.50	125.63	132.71	152.71

1.8 Some issues

NP-hard problem.

Setup:

- gurobipy: parameters
  - WarmStart: (1) Bottom-up, (2) All retained, (3) Relaxed problem `ifelse(z >= 0.001, 1, 0)`.
  - TimeLimit = 600s.
  - MIPGap $= |z_P - z_D| \,/\, |z_P| = 0.01$ for large hierarchy (default value is 10^{-4}), where z_P is the primal objective bound (i.e., the incumbent objective value, which is the upper bound for minimization problems), and z_D is the dual objective bound (i.e., the lower bound for minimization problems),
  - MIPFocus = 3: when the best objective bound is moving very slowly (or not at all), this focus more on the bound.
  - Cuts = 2: aggressive cut generation.
- Bound of variables
  - should be as tight as possible to speed up computation.
  - $\check{e} = \hat{y} - \tilde{y}: \quad [-|y|_{\max}, |y|_{\max}]$.
  - $\mathcal{M}_k$ for each element in G: $[-G^{\text{bench}}_{\max} - 1, G^{\text{bench}}_{\max} + 1]$.
  - $\mathcal{M}$ for sum of absolute values of each column: $[-n_b, n_b]$.

2 Group lasso

2.1 Out-of-sample based method

2.1.1 Group lasso with the unbiasedness constraint

The best-subset selection method is a NP-hard problem, which is computationally intensive.

Instead of involving an ℓ_0 penalty, we consider the following ℓ_1**-regularized regression problem** of the following form to achieve selection in hierarchical forecasting:

$$\begin{aligned}\min_{G} \quad & \frac{1}{2}\left(\hat{y} - \left(\hat{y}' \otimes S\right)\operatorname{vec}(G)\right)' W^{-1}\left(\hat{y} - \left(\hat{y}' \otimes S\right)\operatorname{vec}(G)\right) \\ & \quad + \lambda_1 \sum_{j=1}^{n} w_j \left\|G_{\cdot j}\right\|_2 \\ \text{s.t.} \quad & GS = I_{n_b},\end{aligned} \tag{5}$$

where $\lambda_1 \geqslant 0$ is a tuning parameter, the w_j terms account for the varying group sizes. The ℓ_1-norm penalty induces sparsity in the solution. By introducing such a penalty, group lasso achieves sparse selection not of individual covariates but rather their groups. The target problem is essentially **a group lasso problem with the unbiasedness constraint**.

2.1.2 Second-order cone program

We propose Second-order Cone Program (SOCP) formulations to solve Equation 5.

$$\begin{aligned}\min_{G,z,\check{e},g^{+}} \; & \frac{1}{2}\check{e}' W_h^{-1}\check{e} + \lambda_1 \sum_{j=1}^{n} w_j c_j \\ \text{s.t.} \quad & \hat{y}_h - \left(\hat{y}_h' \otimes S\right)\operatorname{vec}(G) = \check{e} \quad \cdots (C1) \\ & c_j = \sqrt{\sum_{i=1}^{n_b} g_{i+(j-1)n_b}^2}, \quad j \in [n] \quad \cdots (C2) \\ & GS = I_{n_b} \Leftrightarrow \left(S' \otimes I_{n_b}\right)\operatorname{vec}(G) = \operatorname{vec}\left(I_{n_b}\right) \quad \cdots (C3)\end{aligned} \tag{6}$$

where constraint (C2) is a second-order cone.

2.1.3 Some issues

Problem: about w_j and (C3)

In a hierarchy setting, we can consider $w_j = 1$ as each group has the same size. After experiments on simulation data and tourism data, it shows that:

- it can **hardly** shrink some columns of G to 0 when including **the unbiasedness constraint**.
- if we remove the unbiasedness constraint, it frequently gives a **top-down** G or a G with only few number of non-zero columns, and the performance is very **poor and unstable**, especially for longer horizons.

Reason: penalty term $\lambda_1 \sum_{j=1}^{n} w_j \left\|G_{\cdot j}\right\|_2$.

- For a given j, if $G_{\cdot j}$ is zeroed out, then other columns of G tend to be changed, which may have larger ℓ_2-norm values.
- Top level (first column) tends to have smaller ℓ_2-norm when other columns are zeroed out.

Strategy:

- Consider a more flexible group lasso by putting different penalty weights w_j on each group, e.g., $w_j = 1/\left\|G_{\cdot j}^{\text{bench}}\right\|_2$, and also include the unbiasedness constraint.
- λ_1 sequence: $\lambda_{\max} = 10^{k+1}, \quad \lambda_{\min} = 10^{k-4}$ and construct a sequence of nlambda values decreasing from $\lambda_{\max}$ to $\lambda_{\min}$ on the log scale.

2.2 In-sample based method

2.2.1 Empirical group lasso

Here, we consider using the in-sample residuals to formulate the problem. Let Y denote $N \times n$ matrix of historical data of all the time series in the structure, and $\widehat{Y}$ denote the matrix of in-sample 1-step-ahead forecasts of all the time series, where N is the number of historical observations for each series, and n is the number of time series in the hierarchy of interest.

Assuming that **the series in the structure are jointly weakly stationary**, the minimization problem can be given by:

$$
\begin{aligned}
&\min_{G} \quad \frac{1}{2Nn} \left\|Y - \hat{Y} G' S'\right\|_F^2 + \lambda_1 \sum_{j=1}^{n} w_j \left\|G_{\cdot j}\right\|_2 \\
&\quad \Downarrow \\
&\min_{G} \quad \frac{1}{2Nn} \left\|\operatorname{vec}(Y) - (S \otimes \hat{Y}) \operatorname{vec}\left(G'\right)\right\|_2^2 + \lambda_1 \sum_{j=1}^{n} w_j \left\|G_{\cdot j}\right\|_2,
\end{aligned}
$$

After reformulation, it reduced to a standard group lasso problem with $\operatorname{vec}(Y)$ as dependent variable and $S \otimes \hat{Y}$ as design matrix.

Strategy:

- Consider putting different penalty weights w_j on each group, e.g., $w_j = 1/\left\|G_{\cdot j}^{\text{OLS}}\right\|_2$.
- Use the `gglasso` package, specify `intercept = FALSE`, `pf = w`, `lambda.factor = 1e-05`, `eps = 1e-04`, `foldid` to ensure each fold contains t number of observations from each variable (time series).

2.3 Simulation results

2.3.1 Scenario 0: ETS

- ETS models are used to generate base forecasts. Table 10, Table 11, and Figure 5.

Table 10: Out-of-sample average RMSE results in Scenario 0.

Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	9.61	10.73	12.59	15.58	6.33	7.26	8.61	10.83	4.20	4.92	5.93	7.52	5.58	6.41	7.65	9.62
BU	9.51	10.78	12.67	15.68	6.32	7.25	8.62	10.83	4.20	4.92	5.93	7.52	5.56	6.42	7.66	9.63
OLS	9.54	10.71	12.59	15.59	6.33	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.57	6.40	7.63	9.60
OLS-Lasso	9.52	10.72	12.60	15.59	6.31	7.24	8.60	10.81	4.20	4.91	5.92	7.51	5.56	6.41	7.64	9.61
WLSs	9.52	10.71	12.60	15.60	6.32	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
WLSs-Lasso	9.51	10.73	12.61	15.61	6.31	7.24	8.60	10.81	4.20	4.91	5.92	7.51	5.56	6.41	7.64	9.61
WLSv	9.52	10.72	12.60	15.61	6.32	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
WLSv-Lasso	9.51	10.73	12.61	15.61	6.31	7.24	8.60	10.81	4.20	4.91	5.92	7.51	5.56	6.41	7.64	9.61
MinT	9.54	10.74	12.62	15.62	6.31	7.25	8.61	10.82	4.22	4.92	5.93	7.52	5.58	6.42	7.65	9.62
MinT-Lasso	9.54	10.74	12.62	15.62	6.31	7.25	8.61	10.82	4.22	4.92	5.93	7.52	5.58	6.42	7.65	9.62
MinTs	9.52	10.72	12.60	15.60	6.31	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
MinTs-Lasso	9.52	10.71	12.60	15.59	6.31	7.23	8.59	10.80	4.20	4.91	5.92	7.51	5.56	6.40	7.64	9.61
ELasso	9.78	11.02	12.89	15.82	6.48	7.45	8.79	10.96	4.33	5.06	6.05	7.61	5.73	6.60	7.81	9.74

Table 11: Out-of-sample average MASE results in Scenario 0.

Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	0.91	0.87	0.99	1.20	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	0.90	0.88	1.02	1.27
BU	0.90	0.88	1.00	1.21	0.90	0.88	1.01	1.25	0.89	0.88	1.04	1.30	0.89	0.88	1.03	1.27
OLS	0.91	0.87	0.99	1.20	0.90	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.90	0.88	1.02	1.27
OLS-Lasso	0.91	0.87	0.99	1.20	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
WLSs	0.91	0.87	0.99	1.21	0.90	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
WLSs-Lasso	0.90	0.87	0.99	1.21	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
WLSv	0.91	0.87	0.99	1.21	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
WLSv-Lasso	0.90	0.87	0.99	1.21	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
MinT	0.91	0.87	0.99	1.21	0.89	0.88	1.01	1.25	0.89	0.89	1.04	1.29	0.90	0.88	1.02	1.27
MinT-Lasso	0.91	0.87	0.99	1.21	0.89	0.88	1.01	1.25	0.89	0.89	1.04	1.29	0.90	0.88	1.02	1.27
MinTs	0.91	0.87	0.99	1.20	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
MinTs-Lasso	0.91	0.87	0.99	1.20	0.89	0.87	1.01	1.25	0.89	0.88	1.04	1.29	0.89	0.88	1.02	1.27
ELasso	0.93	0.89	1.01	1.22	0.92	0.90	1.03	1.27	0.92	0.91	1.06	1.31	0.92	0.90	1.05	1.29

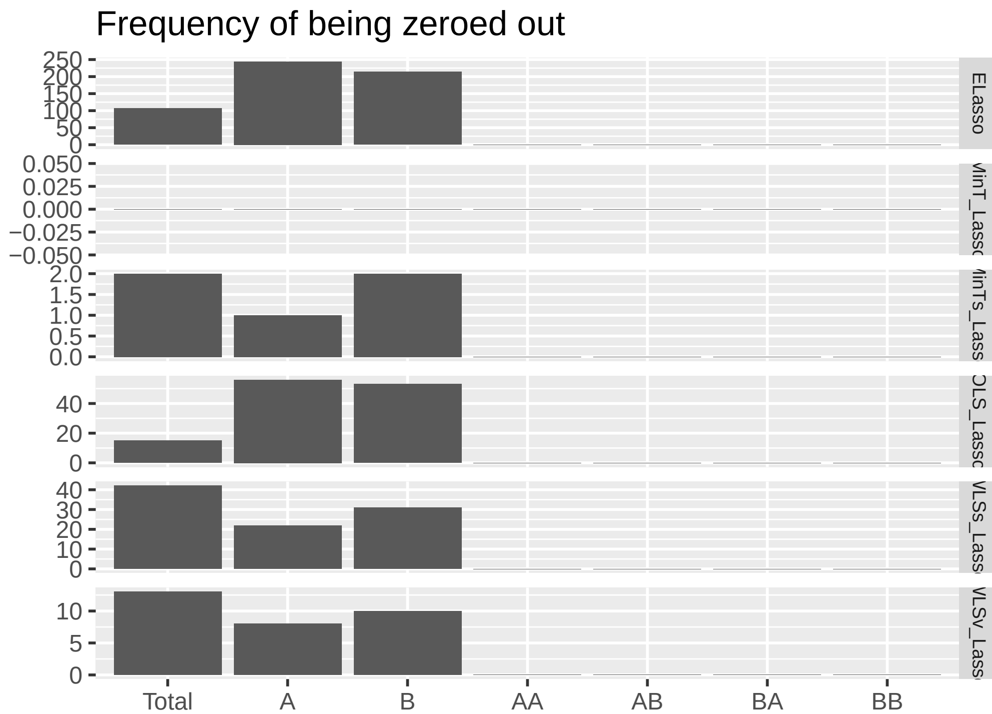

Figure 5: Frequency of the base forecasts being removed from reconciliation in Scenario 0.

2.3.2 Scenario 1: D-AA

- Base forecasts (and also fitted values) of **series AA** multiplied by 1.5 to achieve deterioration. **?@tbl-lasso-rmse-s1**, **?@tbl-lasso-mase-s1**, and **?@fig-lasso-s1**.

2.3.3 Scenario 2: D-A

- Base forecasts (and also fitted values) of **series A** multiplied by 1.5 to achieve deterioration. **?@tbl-lasso-rmse-s2**, **?@tbl-lasso-mase-s2**, and **?@fig-lasso-s2**.

2.3.4 Scenario 3: D-Total

- Base forecasts (and also fitted values) of **series Total** multiplied by 1.5 to achieve deterioration. **?@tbl-lasso-rmse-s3**, **?@tbl-lasso-mase-s3** and **?@fig-lasso-s3**.

2.4 Tourism data results

- OLS_Lasso: $\lambda_1 =, \sum z =$
- WLSs_Lasso: $\lambda_1 =, \sum z =$
- WLSv_Lasso: $\lambda_1 =, \sum z =$
- MinTs_Lasso: $\lambda_1 =, \sum z =$
- ELasso: $\lambda_1 =, \sum z =$

3 Intuitive method

4 Further issues

- Parallel issue: We cannot export **reticulate** `python.builtin.module` objects from one R process to another. They are designed to only work within the same R process they're created.
- HPC: gurobi license.
- Simulation: include structural break.
- Large hierarchy:
  - Candidate lambda sequence
  - Sub hierarchy + Voting
- Theoretical aspects
- Grouped time series

Hastie, Trevor, Robert Tibshirani, and Ryan Tibshirani. 2020. "Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons." *Statistical Science* 35 (4). https://doi.org/10.1214/19-sts733.

Mazumder, Rahul, Peter Radchenko, and Antoine Dedieu. 2023. "Subset Selection with Shrinkage: Sparse Linear Modeling When the SNR Is Low." *Operations Research* 71 (1): 129–47. https://doi.org/10.1287/opre.2022.2276.

Table 12: Out-of-sample average RMSE results in Tourism data.

	Top				Sate				Zone				Region				Average			
Method	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12	h=1	1-4	1-8	1-12
Base	1158.16	716.62	1279.50	1907.61	452.68	323.31	349.92	424.77	165.52	163.62	160.69	179.71	100.80	89.43	88.25	94.11	148.26	127.87	133.11	152.12
BU	2189.97	1667.97	1962.73	2708.64	431.95	356.46	409.53	508.37	167.29	159.70	161.40	181.55	100.80	89.43	88.25	94.11	156.68	137.58	143.19	165.06
OLS	1103.20	714.05	1286.29	1935.26	438.89	310.65	344.21	418.35	162.12	156.78	151.75	166.20	101.79	89.07	86.56	91.13	146.75	125.14	129.48	146.64
OLS-Lasso	1243.79	765.41	1308.09	1943.16	407.88	493.15	708.45	769.78	189.08	207.66	281.62	301.80	121.21	126.83	157.22	162.21	165.91	175.34	232.61	250.53
WLSs	1448.95	1111.95	1546.09	2271.59	381.21	307.05	362.27	451.23	155.70	154.71	153.10	170.80	100.60	88.72	86.85	92.07	143.85	127.76	133.48	153.51
WLSs-Lasso	1528.83	936.10	1413.26	2066.35	379.84	458.45	670.58	737.36	157.16	213.72	292.62	312.34	100.94	120.96	154.88	160.62	145.07	172.15	232.24	251.06
WLSv	1600.65	1262.35	1657.96	2395.97	374.05	313.35	374.44	466.85	157.23	156.62	155.64	173.97	96.61	88.01	86.66	92.16	142.40	129.49	135.74	156.44
WLSv-Lasso	1601.70	981.23	1442.34	2098.94	374.03	449.42	660.14	727.72	157.24	216.40	296.06	315.25	96.61	120.25	154.48	160.29	142.41	172.15	232.41	251.23
MinTs	1397.23	1101.06	1555.74	2270.36	352.16	300.15	362.01	451.41	145.46	152.82	152.55	170.11	95.41	87.07	85.82	91.15	135.50	125.63	132.71	152.71
MinTs-Lasso	1399.29	857.21	1363.67	2009.49	352.16	438.92	655.13	720.45	145.49	213.37	295.29	314.52	95.41	119.77	154.47	160.30	135.53	169.31	231.19	249.80
ELasso	2189.97	1667.97	1962.73	2708.64	431.95	356.46	409.53	508.37	167.29	159.70	161.40	181.55	100.80	89.43	88.25	94.11	156.68	137.58	143.19	165.06