Forecast reconciliation with subset selection

Table of contents

1 Group best-subset selection **2**
- 1.1 MinT reconciliation . . . 2
- 1.2 Best-subset selection . . . 2
- 1.3 Group best-subset selection with ridge regularization . . . 3
- 1.4 Mixed integer program . . . 3
- 1.5 Hyperparameter . . . 4
- 1.6 Simulation results . . . 5
  - 1.6.1 Data simulation . . . 5
  - 1.6.2 Scenarios: . . . 6
- 1.7 Tourism data results . . . 6
  - 1.7.1 Data description . . . 6
- 1.8 Some issues . . . 6

2 Group lasso **7**
- 2.1 Out-of-sample based method . . . 7
  - 2.1.1 Group lasso with the unbiasedness constraint . . . 7
  - 2.1.2 Second-order cone program . . . 8
  - 2.1.3 Strategy & hyperparameter . . . 8
- 2.2 In-sample based method . . . 10
  - 2.2.1 Empirical group lasso . . . 10
  - 2.2.2 Strategy & hyperparameter . . . 10

3 Intuitive method **11**
- 3.1 Methodology . . . 11
- 3.2 Hyperparameter . . . 12

4 More on the unbiasedness constraint **12**

5 Results **13**
- 5.1 Simulation data . . . 13

5.2 Tourism data 16

6 Further issues **18**

7 References **21**

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:

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

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:

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

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

1. **Former strategy**

- $\lambda_0 = \{0, 10^{k-3}, 10^{k-2}, 10^{k-1}, 10^k, 10^{k+1}\}$, 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.

2. **New strategy**

- $\lambda_{0\max} = \frac{1}{2}\left(\hat{y}_h - \tilde{y}_h^{\text{MinT}}\right)' W_h^{-1}\left(\hat{y}_h - \tilde{y}_h^{\text{MinT}}\right)$, $\lambda_{0\min} = 0.0001\lambda_{0\max}$, We compute solutions over a grid of $k-1$ values between $\lambda_{0\min}$ and $\lambda_{0\max}$, where $\lambda_{0,j} = \lambda_{0\max}\left(\lambda_{0\min}/\lambda_{0\max}\right)^{j/(k-1)}$ for $j = 0, \ldots, k-1$. Thus, $\lambda_0 = \{0, \lambda_{0,0}, \ldots, \lambda_{0,k-1}\}$. In our implementation, the default value for k is 20.

- $\lambda_2 = \{0, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$

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,

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

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 Scenarios:

- Scenario 0: ETS
  - ETS models are used to generate base forecasts.
- Scenario I: D-AA
  - Base forecasts (and also fitted values) of **series AA** multiplied by 1.5 to achieve deterioration.
- Scenario II: D-A
  - Base forecasts (and also fitted values) of **series A** multiplied by 1.5 to achieve deterioration.
- Scenario III: D-Total
  - Base forecasts (and also fitted values) of **series Total** multiplied by 1.5 to achieve deterioration.

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.

1.8 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: $[-|G|_{\max}^{\text{bench}} - 1, |G|_{\max}^{\text{bench}} + 1]$.
  - $\mathcal{M}$ 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:

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

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 Strategy & hyperparameter

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.

Penalty weights

- 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.

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: $[-|G|_{\max}^{\text{bench}} - 1, |G|_{\max}^{\text{bench}} + 1]$.
- $\mathcal{M}$ for sum of absolute values of each column: $[-n_b, n_b]$.

λ Sequence

1. **Former strategy**

- $\lambda_1 = \{0, 10^{k-3}, 10^{k-2}, 10^{k-1}, 10^k, 10^{k+1}\}$, where k is the number of digits before the decimal point for $\frac{1}{2}\left(\hat{y}_h - \tilde{y}_h^{\text{MinT}}\right)' W_h^{-1}\left(\hat{y}_h - \tilde{y}_h^{\text{MinT}}\right)$

2. **New strategy**

The method used in (Yang and Zou 2014) to decide a sequence of λ values for group-lasso penalize learning problem is shown below.

The objective function is

$$L(\beta \mid \mathbf{D}) + \lambda \sum_{j=1}^{n} w_j \left\| \beta^{(j)} \right\|_2 .$$

We define $\lambda^{[1]}$ as the smallest λ value such that all predictors (without intercept) have zero coefficients. Then the solution at $\lambda^{[1]}$ is $\hat{\beta}^{[1]} = 0$ as the null model estimates. Our strategy is to select a minimum value $\lambda_{\min}$, and construct a sequence of K values of λ decreasing from λ_{max} to $\lambda_{\min}$ on the log scale.

$$\lambda_{\max} = \lambda^{[1]} = \max_{j=1,\ldots,n} \left\| \left[\nabla L\left(\hat{\beta}^{[1]} \mid \mathbf{D}\right) \right]^{(j)} \right\|_2 / w_j, \quad w_j \neq 0$$

For the out-of-sample based group lasso method, we ignore the unbiasedness constraint when we decide $\lambda_{\max}$. As W^{-1} is symmetric,

$$\lambda_{\max} = \max_{j=1,\ldots,n} \left\| - \left(\left(\hat{y}' \otimes S \right)_{\cdot cj} \right)' W^{-1} \hat{y} \right\|_2 / w_j,$$

where cj is the column indices of $\left(\hat{y}' \otimes S\right)$ corresponding to the jth column of G.

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 $\hat{Y}$ 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:

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

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

2.2.2 Strategy & hyperparameter

- We can use the `gglasso` package, specifically we set `intercept = FALSE`, `pf = w`, `lambda.factor = 1e-05`, `eps = 1e-04`, and `foldid` to ensure each fold contains t number of observations from each variable (time series). It's very slow when we set penalty factor by setting the `pf` parameter.
- Thus, we consider proposing SOCP formulations to solve it, as in the out-of-sample based method.

Penalty weights

- Consider putting different penalty weights w_j on each group, e.g., $w_j = 1/\left\| G_{\cdot j}^{\mathrm{OLS}} \right\|_2, j = 1, 2, \ldots, n$.

λ Sequence

1. **Former strategy**

λ_1 sequence: $\lambda_1 = \{0, 10^{k-3}, 10^{k-2}, 10^{k-1}, 10^k, 10^{k+1}\}$, where k is the number of digits before the decimal point for $\frac{1}{2N^{\text{train}}} \left\| Y^{\text{train}} - \hat{Y}^{\text{train}} G'^{\text{MinT}} S' \right\|_F^2$.

2. **New strategy**

For the in-sample based group lasso method,

$$\lambda_{\max} = \max_{j=1,\ldots,n} \left\| -\frac{1}{N} \left(\left(S \otimes \hat{Y} \right)_{\cdot cj} \right)' \text{vec}(Y) \right\|_2 / w_j$$

where cj is the column indices of $(S \otimes \hat{Y})$ corresponding to the jth column of G.

3 Intuitive method

3.1 Methodology

Let $\bar{S} = AS$, where $A = \text{diag}(z_i)$ is a **diagonal matrix** with $z_i \in \{0, 1\}$. Then $\bar{G} = (S'A'W^{-1}AS)^{-1}S'A'W^{-1}$.

The problem of estimating the whole G reduces to estimating an appropriate A and then obtaining $\bar{G}$.

Therefore, the problem can be reduced to a minimization problem as follows:

$$\begin{aligned} \min_{A} \quad & \frac{1}{2} \left(\hat{y} - S\bar{G}\hat{y} \right)' W^{-1} \left(\hat{y} - S\bar{G}\hat{y} \right) + \lambda_0 \sum_{j=1}^{n} z_i \\ \text{s.t.} \quad & \bar{G} = (S'A'W^{-1}AS)^{-1}S'A'W^{-1} \\ & \bar{G}S = I \end{aligned}$$

The main problem with the calculation is that we need to ensure that $(S'A'W^{-1}AS)$ is invertible. In Gurobi, we can formulate the problem as follows.

$$
\begin{aligned}
\min_{A,\bar{G},C} \quad & \frac{1}{2}\left(\hat{y}-\left(\hat{y}'\otimes S\right)\mathrm{vec}(\bar{G})\right)' W^{-1}\left(\hat{y}-\left(\hat{y}'\otimes S\right)\mathrm{vec}(\bar{G})\right)+\lambda_0\sum_{j=1}^{n} z_i \\
\text{s.t.} \quad & C(S'A'W^{-1}AS)=I \\
& \bar{G}=CS'A'W^{-1} \\
& \bar{G}S=I
\end{aligned}
$$

To be able to use Gurobi, we rewrite the problem (constraints) as follows.

$$
\begin{aligned}
\min_{A,\bar{G},C} \quad & \frac{1}{2}\left(\hat{y}-\left(\hat{y}'\otimes S\right)\mathrm{vec}(\bar{G})\right)' W^{-1}\left(\hat{y}-\left(\hat{y}'\otimes S\right)\mathrm{vec}(\bar{G})\right)+\lambda_0\sum_{j=1}^{n} z_i \\
\text{s.t.} \quad & \bar{G}AS=I \\
& \bar{G}=CS'A'W^{-1} \\
& \bar{G}S=I
\end{aligned}
$$

3.2 Hyperparameter

- $\lambda_{0\max} = \frac{1}{2n_b}\left(\hat{y}_h-\tilde{y}_h^{\mathrm{MinT}}\right)' W_h^{-1}\left(\hat{y}_h-\tilde{y}_h^{\mathrm{MinT}}\right)$, $\lambda_{0\min} = 0.0001\lambda_{0\max}$, We compute solutions over a grid of $k-1$ values between $\lambda_{0\min}$ and $\lambda_{0\max}$, where $\lambda_{0,j}=\lambda_{0\max}\left(\lambda_{0\min}/\lambda_{0\max}\right)^{j/(k-1)}$ for $j=0,\dots,k-1$. Thus, $\lambda_0=\{0,\lambda_{0,0},\dots,\lambda_{0,k-1}\}$. In our implementation, the default value for k is 20.

4 More on the unbiasedness constraint

The unbiasedness constraint is

$$GS=I$$

and our focus is to estimate G by minimizing a loss function with several constraints.

According to the unbiasedness constraint, we have

1. **Rank — Number of non-zero columns**

   Conclusion: $\mathrm{rank}(G)\geq n_b$ which means that the **number** of non-zero columns of the estimated G should be at least n_b.

   Proof: $\min(\mathrm{rank}(G),\mathrm{rank}(S))\geq\mathrm{rank}(I)=n_b$

2. **Hierarchical structure — Location of non-zero columns**

   Let $G_{\cdot\mathbb{S}} \in \mathbb{R}^{n_b \times |\mathbb{S}|}$ denote the submatrix of G whose columns are indexed by a set $\mathbb{S}$ (and when $\mathbb{S} = \{j\}$, we simply use G_j), and $G_{\mathbb{S}\cdot} \in \mathbb{R}^{|\mathbb{S}| \times n}$ denote the submatrix of G whose rows are indexed by a set $\mathbb{S}$ (and when $\mathbb{S} = \{i\}$, we simply use $G_{i\cdot}$),

   Conclusion: If the set $\mathbb{S}$ involves the indices of non-zero columns of the estimated G, then $\text{rank}(S_{\mathbb{S}\cdot}) = n_b$ which means that we should make sure that the whole hierarchy is still **revertible** after removing the nodes corresponding to zero columns of G.

   Proof: If the set $\mathbb{S}$ involves the indices of non-zero columns of the estimated G, then $GS = G_{\cdot\mathbb{S}} S_{\mathbb{S}\cdot}$ and $\min(\text{rank}\,(G_{\cdot\mathbb{S}}), \text{rank}(S_{\mathbb{S}\cdot})) \geq \text{rank}(I) = n_b$.

   Recall that $\text{rank}(S_{\mathbb{S}\cdot}) \leq n_b$ as S has n_b columns. Hence, $\text{rank}(S_{\mathbb{S}\cdot}) = n_b$.

5 Results

5.1 Simulation data

Table 1: Out-of-sample forecast performance (average RMSE) for the simulation data.

	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	9.6	10.7	12.6	15.6	6.3	7.3	8.6	10.8	4.2	4.9	5.9	7.5	5.6	6.4	7.6	9.6
BU	-1.0	0.4	0.6	0.7	-0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	-0.3	0.1	0.2	0.2
OLS	-0.7	**-0.2**	**0.0**	**0.0**	-0.1	**-0.3**	**-0.2**	**-0.3**	0.1	**-0.2**	**-0.2**	**-0.1**	-0.2	**-0.2**	**-0.2**	**-0.1**
OLS-subset	**-0.8**	0.2	0.3	0.4	**-0.2**	-0.1	0.0	-0.1	0.1	0.1	0.0	0.1	-0.2	0.0	0.1	0.1
OLS-intuitive	**-0.9**	-0.1	0.1	0.2	**-0.2**	**-0.3**	-0.1	-0.1	0.2	0.0	0.0	0.0	-0.2	-0.1	0.0	0.0
OLS-lasso	**-1.3**	-0.1	0.3	0.4	**-0.5**	**-0.3**	-0.1	-0.1	**-0.1**	-0.1	-0.1	**-0.1**	**-0.5**	**-0.2**	0.0	0.0
WLSs	-0.9	-0.1	**0.0**	0.2	-0.3	**-0.3**	**-0.2**	-0.2	0.0	**-0.2**	**-0.2**	**-0.1**	-0.3	**-0.2**	-0.1	**-0.1**
WLSs-subset	**-1.0**	0.1	0.3	0.3	-0.2	-0.2	0.0	-0.1	0.1	0.0	-0.1	0.0	-0.3	-0.1	0.0	0.0
WLSs-intuitive	**-1.0**	-0.1	0.1	0.3	-0.3	**-0.3**	-0.1	-0.1	0.1	-0.1	-0.1	0.0	-0.3	**-0.2**	-0.1	0.0
WLSs-lasso	**-1.3**	0.0	0.3	0.5	**-0.5**	-0.2	0.0	-0.1	**-0.1**	-0.1	-0.1	0.0	**-0.5**	-0.1	0.0	0.1
WLSv	-0.9	-0.1	0.1	0.2	-0.3	**-0.3**	**-0.2**	-0.2	0.0	**-0.2**	**-0.2**	**-0.1**	-0.3	**-0.2**	-0.1	**-0.1**
WLSv-subset	-0.9	0.2	0.4	0.5	-0.3	-0.1	0.1	0.0	0.0	0.0	0.0	0.1	-0.3	0.0	0.1	0.2
WLSv-intuitive	**-1.0**	0.0	0.2	0.3	-0.3	-0.2	-0.1	-0.1	0.0	0.0	0.0	0.0	**-0.4**	-0.1	0.0	0.0
WLSv-lasso	**-1.3**	0.0	0.3	0.5	**-0.5**	-0.2	0.0	-0.1	**-0.1**	-0.1	-0.1	0.0	**-0.5**	-0.1	0.0	0.1
MinT	-0.7	0.1	0.2	0.2	-0.3	-0.1	0.0	-0.1	0.4	0.1	0.0	**-0.1**	-0.1	0.1	0.1	0.0
MinT-subset	-0.7	0.3	0.5	0.6	-0.2	0.1	0.2	0.1	**0.3**	0.2	0.1	0.1	-0.1	0.2	0.2	0.2
MinT-intuitive	-0.7	0.1	0.2	0.2	-0.3	-0.1	0.0	-0.1	0.4	0.1	0.0	**-0.1**	-0.1	0.1	0.1	0.0
MinT-lasso	**-1.3**	**-0.1**	0.2	0.3	**-0.6**	**-0.2**	0.0	-0.1	**0.3**	**0.0**	0.0	0.0	**-0.4**	**-0.1**	**0.0**	0.0
MinTs	-0.9	-0.1	0.1	0.1	-0.4	**-0.3**	**-0.2**	**-0.3**	0.1	-0.1	**-0.2**	**-0.1**	-0.3	**-0.2**	-0.1	**-0.1**
MinTs-subset	**-1.0**	0.1	0.2	0.4	-0.4	-0.2	-0.1	-0.1	**0.0**	0.0	0.0	0.0	**-0.4**	-0.1	0.0	0.1
MinTs-intuitive	-0.9	-0.1	0.1	0.1	-0.4	**-0.3**	**-0.2**	**-0.3**	0.1	-0.1	**-0.2**	**-0.1**	-0.3	**-0.2**	-0.1	**-0.1**
MinTs-lasso	**-1.4**	-0.1	0.2	0.4	**-0.6**	**-0.3**	-0.1	-0.1	**-0.1**	-0.1	-0.1	**-0.1**	**-0.6**	**-0.2**	0.0	0.0
Elasso	1.6	2.8	2.4	1.6	2.2	2.8	2.3	1.3	3.2	3.2	2.1	1.2	2.5	3.0	2.3	1.4

Table 2: Out-of-sample forecast performance (average RMSE) for the simulation data in Scenario I (Base forecasts and also fitted values of series AA are deteriorated).

	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	9.6	10.7	12.6	15.6	6.3	7.3	8.6	10.8	6.4	7.5	8.3	9.8	6.8	7.9	9.0	10.9
BU	57.8	68.5	53.7	38.9	58.2	61.8	48.1	34.4	0.0	0.0	0.0	0.0	27.0	29.6	23.8	17.7
OLS	0.6	2.2	1.8	1.4	7.1	6.4	4.6	3.1	-7.6	-8.6	-8.2	-7.3	-2.1	-2.5	-2.7	-2.6
OLS-subset	0.6	**1.8**	**1.5**	**1.3**	7.2	**5.2**	**3.8**	**2.6**	**-8.3**	**-12.9**	**-11.6**	**-9.9**	**-2.4**	**-5.2**	**-4.8**	**-4.1**
OLS-intuitive	0.8	2.6	2.1	1.8	7.5	**6.1**	**4.4**	**3.0**	**-9.0**	**-12.8**	**-11.6**	**-9.9**	**-2.7**	**-4.8**	**-4.5**	**-3.8**
OLS-lasso	0.6	2.2	1.8	1.6	7.4	6.7	4.8	3.2	-7.6	-8.5	-8.1	-7.2	-2.0	-2.4	-2.6	-2.5
WLSs	7.3	10.6	8.1	5.9	15.6	16.0	11.8	8.0	-6.9	-7.8	-7.4	-6.4	1.9	2.0	1.0	0.2
WLSs-subset	**5.0**	**5.7**	**4.6**	**3.6**	**12.3**	**10.0**	**7.5**	**5.2**	**-7.6**	**-10.5**	**-9.6**	**-8.2**	**0.2**	**-2.0**	**-2.1**	**-2.0**
WLSs-intuitive	**7.1**	**9.2**	**7.1**	**5.2**	16.5	**15.5**	**11.5**	**7.9**	-6.8	**-9.2**	**-8.4**	**-7.3**	2.1	**0.9**	**0.1**	**-0.4**
WLSs-lasso	7.3	**10.3**	**8.0**	5.9	15.7	16.1	11.8	8.1	**-7.0**	-7.8	-7.3	-6.4	1.9	2.0	1.0	0.2
WLSv	1.0	2.9	2.3	1.9	4.5	4.3	3.2	2.1	-25.8	-26.4	-22.7	-18.3	-12.4	-12.6	-10.7	-8.4
WLSv-subset	**-1.0**	**0.3**	**0.4**	**0.5**	**0.6**	**0.6**	**0.5**	**0.3**	**-32.3**	**-32.2**	**-27.3**	**-21.7**	**-17.3**	**-17.3**	**-14.2**	**-10.9**
WLSv-intuitive	**-0.5**	**0.2**	**0.3**	**0.5**	**0.9**	**0.7**	**0.5**	**0.3**	**-32.3**	**-32.3**	**-27.4**	**-21.7**	**-17.1**	**-17.3**	**-14.2**	**-10.9**
WLSv-lasso	**0.4**	**1.5**	**1.5**	**1.4**	**3.0**	**2.5**	**2.0**	**1.3**	**-28.5**	**-29.2**	**-24.9**	**-19.9**	**-14.4**	**-14.9**	**-12.3**	**-9.5**
MinT	-0.4	0.7	0.9	0.6	0.7	0.7	0.6	0.3	-32.9	-33.4	-28.3	-22.5	-17.5	-17.8	-14.6	-11.3
MinT-subset	**-0.6**	0.7	**0.8**	0.7	**0.6**	0.8	0.6	0.3	**-33.0**	-33.1	-28.0	-22.3	**-17.6**	-17.6	-14.5	-11.2
MinT-intuitive	-0.4	0.7	0.9	0.6	0.7	0.7	0.6	0.3	-32.9	-33.4	-28.3	-22.5	-17.5	-17.8	-14.6	-11.3
MinT-lasso	**-0.7**	**0.3**	**0.6**	**0.4**	**0.3**	**0.4**	**0.4**	**0.1**	**-33.2**	**-33.7**	**-28.5**	**-22.6**	**-17.8**	**-18.1**	**-14.8**	**-11.4**
MinTs	-0.9	0.6	0.7	0.5	0.6	0.6	0.5	0.2	-32.9	-33.5	-28.3	-22.5	-17.6	-17.9	-14.6	-11.3
MinTs-subset	-0.7	0.9	1.1	1.0	0.7	0.8	0.7	0.4	**-33.0**	-33.1	-27.9	-22.2	-17.6	-17.5	-14.3	-11.0
MinTs-intuitive	-0.9	0.6	0.7	0.5	0.6	0.6	0.5	0.2	-32.9	-33.5	-28.3	-22.5	-17.6	-17.9	-14.6	-11.3
MinTs-lasso	-0.9	**0.4**	**0.6**	0.5	0.6	**0.4**	**0.4**	**0.1**	**-33.2**	**-33.6**	**-28.4**	**-22.6**	**-17.7**	**-18.0**	**-14.8**	**-11.4**
Elasso	1.5	2.8	2.4	1.7	2.1	2.8	2.3	1.3	-32.1	-32.2	-27.4	-21.9	-16.3	-16.2	-13.3	-10.5

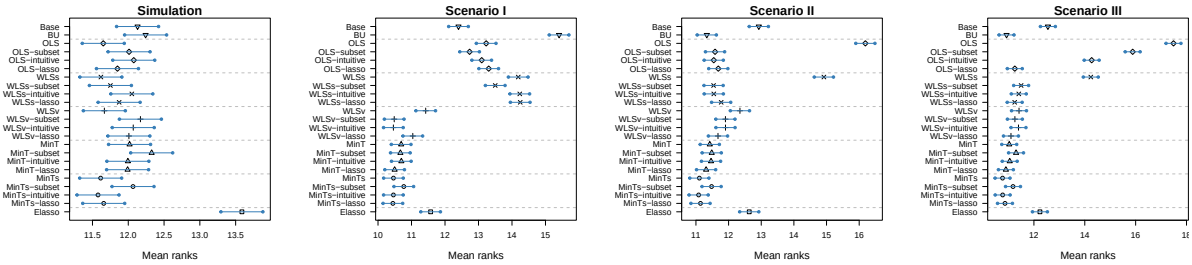


Figure 1: MCB test conducted on the methods examined using the simulation data, the ranks are computed considering average accuracy for all the time series in 500 hierarchies, i.e. a total of 7 series $\times$ 500 hierarchies = 3500 instances.

Table 3: Out-of-sample forecast performance (average RMSE) for the simulation data in Scenario II (Base forecasts and also fitted values of series A are deteriorated).

	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	9.6	10.7	12.6	15.6	12.1	14.4	15.3	17.0	4.2	4.9	5.9	7.5	7.2	8.5	9.6	11.4
BU	**-1.0**	0.4	0.6	0.7	**-47.7**	**-49.6**	**-43.6**	**-36.2**	**0.0**	**0.0**	**0.0**	**0.0**	**-23.0**	**-24.0**	**-19.8**	**-15.3**
OLS	8.5	13.9	10.4	7.6	-28.2	-29.4	-26.7	-23.1	22.9	23.9	17.0	11.3	-4.2	-3.8	-4.2	-4.1
OLS-subset	**-0.5**	**0.5**	**0.6**	**0.7**	**-46.3**	**-49.0**	**-43.2**	**-35.9**	**2.2**	**1.0**	**0.7**	**0.5**	**-21.5**	**-23.4**	**-19.4**	**-15.0**
OLS-intuitive	**-0.5**	**0.5**	**0.6**	**0.6**	**-46.5**	**-49.0**	**-43.2**	**-36.0**	**2.2**	**1.2**	**0.7**	**0.5**	**-21.6**	**-23.4**	**-19.4**	**-15.0**
OLS-lasso	**-0.2**	**1.5**	**1.4**	**1.3**	**-46.9**	**-48.9**	**-43.1**	**-35.8**	**0.9**	**0.8**	**0.5**	**0.3**	**-22.1**	**-23.3**	**-19.3**	**-14.9**
WLSs	12.1	18.6	14.0	10.2	-34.4	-35.1	-31.7	-26.9	15.6	17.0	12.0	8.0	-9.0	-8.0	-7.6	-6.5
WLSs-subset	**-0.1**	**1.2**	**1.1**	**1.1**	**-46.7**	**-48.8**	**-43.1**	**-35.8**	**1.5**	**1.1**	**0.8**	**0.6**	**-21.8**	**-23.2**	**-19.2**	**-14.8**
WLSs-intuitive	**0.0**	**1.2**	**1.0**	**0.9**	**-46.5**	**-48.8**	**-43.1**	**-35.9**	**1.7**	**1.3**	**0.9**	**0.6**	**-21.6**	**-23.1**	**-19.2**	**-14.9**
WLSs-lasso	**-0.1**	**1.5**	**1.5**	**1.3**	**-46.7**	**-48.9**	**-43.1**	**-35.8**	**0.9**	**0.8**	**0.5**	**0.3**	**-22.0**	**-23.2**	**-19.3**	**-14.9**
WLSv	-0.8	2.3	1.8	1.6	-46.3	-47.9	-42.3	-35.2	1.6	1.9	1.2	0.8	-21.7	-22.2	-18.6	-14.4
WLSv-subset	-0.7	**1.3**	**1.4**	**1.4**	**-46.9**	**-48.7**	**-42.9**	**-35.6**	**1.0**	**1.0**	**0.8**	**0.6**	**-22.2**	**-23.1**	**-19.1**	**-14.7**
WLSv-intuitive	-0.4	**1.5**	**1.4**	**1.2**	**-46.9**	**-48.6**	**-42.8**	**-35.6**	**0.9**	**1.2**	**0.9**	**0.7**	**-22.2**	**-23.0**	**-19.0**	**-14.7**
WLSv-lasso	-0.6	**1.3**	**1.3**	**1.3**	**-47.2**	**-48.9**	**-43.0**	**-35.7**	**0.6**	**0.8**	**0.5**	**0.4**	**-22.4**	**-23.3**	**-19.2**	**-14.8**
MinT	0.2	0.5	0.6	0.5	-47.5	-49.4	-43.5	-36.1	1.1	0.5	0.3	0.1	-22.3	-23.7	-19.6	**-15.3**
MinT-subset	**-0.1**	0.8	0.9	0.9	-46.9	-49.1	-43.3	-36.0	1.7	0.9	0.5	0.3	-21.9	-23.4	-19.4	-15.1
MinT-intuitive	0.2	0.5	0.6	0.5	-47.5	-49.4	-43.5	-36.1	1.1	0.5	0.3	0.1	-22.3	-23.7	-19.6	**-15.3**
MinT-lasso	**-0.3**	**0.3**	0.6	0.5	**-47.6**	-49.4	-43.5	-36.1	**0.8**	**0.3**	**0.2**	0.1	**-22.5**	**-23.9**	**-19.7**	**-15.3**
MinTs	-0.3	0.3	**0.4**	**0.4**	-47.6	-49.5	**-43.6**	**-36.2**	0.7	0.2	0.1	**0.0**	-22.6	-23.9	**-19.8**	**-15.3**
MinTs-subset	**-0.8**	0.5	0.8	0.8	-47.2	-49.2	-43.4	-36.0	1.0	0.7	0.4	0.3	-22.3	-23.6	-19.5	-15.1
MinTs-intuitive	-0.3	0.3	**0.4**	**0.4**	-47.6	-49.5	**-43.6**	**-36.2**	0.7	0.2	0.1	**0.0**	-22.6	-23.9	**-19.8**	**-15.3**
MinTs-lasso	**-0.9**	**0.2**	0.5	0.5	**-47.7**	-49.5	**-43.6**	**-36.2**	**0.5**	0.2	0.1	0.1	**-22.8**	**-24.0**	**-19.8**	**-15.3**
Elasso	1.4	2.7	2.4	1.6	-46.4	-48.2	-42.4	-35.4	3.1	3.2	2.1	1.2	-20.9	-21.9	-18.2	-14.3

Table 4: Out-of-sample forecast performance (average RMSE) for the simulation data in Scenario III (Base forecasts and also fitted values of series Total are deteriorated).

	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	25.0	30.3	30.9	32.3	6.3	7.3	8.6	10.8	4.2	4.9	5.9	7.5	7.8	9.2	10.3	12.0
BU	-62.0	**-64.4**	**-59.0**	-51.5	**-0.3**	**0.0**	**0.1**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**-28.5**	**-30.2**	**-25.3**	**-19.8**
OLS	-34.8	-35.5	-33.5	-30.1	45.3	50.6	37.7	25.1	27.7	29.9	21.2	13.7	3.1	3.8	1.6	-0.2
OLS-subset	**-35.3**	**-41.9**	**-39.2**	**-35.0**	**43.9**	**39.5**	**29.5**	**19.6**	**27.1**	**23.6**	**16.8**	**10.9**	**2.4**	**-3.5**	**-4.2**	**-4.5**
OLS-intuitive	**-41.2**	**-49.2**	**-45.5**	**-40.0**	**35.1**	**26.8**	**20.3**	**13.7**	**21.9**	**15.9**	**11.5**	**7.6**	**-4.0**	**-12.2**	**-10.9**	**-9.1**
OLS-lasso	**-61.8**	**-63.6**	**-58.1**	**-50.9**	**0.4**	**1.3**	**1.3**	**0.7**	**0.3**	**0.8**	**0.6**	**0.4**	**-28.2**	**-29.3**	**-24.5**	**-19.2**
WLSs	-50.9	-52.4	-48.7	-43.3	17.6	20.0	14.5	9.3	9.6	11.3	7.7	4.9	-16.3	-16.7	-14.9	-12.5
WLSs-subset	**-61.8**	**-63.6**	**-58.1**	**-50.7**	**0.3**	**1.4**	**1.4**	**0.9**	**0.3**	**0.9**	**0.7**	**0.6**	**-28.2**	**-29.3**	**-24.4**	**-19.0**
WLSs-intuitive	**-61.8**	**-63.8**	**-58.3**	**-50.9**	**0.0**	**1.0**	**1.0**	**0.7**	**0.3**	**0.7**	**0.6**	**0.5**	**-28.3**	**-29.5**	**-24.6**	**-19.2**
WLSs-lasso	**-61.7**	**-63.5**	**-58.0**	**-50.7**	**0.5**	**1.5**	**1.4**	**0.9**	**0.3**	**0.9**	**0.7**	**0.5**	**-28.1**	**-29.2**	**-24.4**	**-19.1**
WLSv	-61.1	-63.4	-58.1	-50.8	1.0	1.7	1.3	0.8	0.7	1.0	0.6	0.4	-27.6	-29.1	-24.5	-19.2
WLSv-subset	**-61.9**	**-63.6**	**-58.2**	**-50.9**	**0.2**	**1.3**	**1.2**	0.8	**0.1**	**0.8**	0.6	0.5	**-28.3**	**-29.3**	-24.5	-19.2
WLSv-intuitive	**-61.8**	**-63.8**	**-58.3**	**-51.0**	**0.0**	**1.1**	**1.1**	**0.6**	**0.1**	**0.6**	**0.5**	0.4	**-28.4**	**-29.5**	**-24.7**	**-19.3**
WLSv-lasso	**-61.8**	**-63.9**	**-58.4**	**-51.1**	**0.2**	**0.9**	**0.9**	**0.5**	**0.1**	**0.5**	**0.4**	**0.3**	**-28.3**	**-29.6**	**-24.8**	**-19.4**
MinT	-62.1	-64.3	-58.9	**-51.6**	-0.2	0.6	0.5	0.2	0.8	0.5	0.3	0.1	-28.3	-29.9	-25.1	**-19.8**
MinT-subset	-61.8	-63.7	-58.2	-50.9	0.4	1.2	1.3	0.8	0.8	1.0	0.7	0.5	-28.0	-29.3	-24.5	-19.2
MinT-intuitive	-62.1	-64.3	-58.9	**-51.6**	-0.2	0.6	0.5	0.2	0.8	0.5	0.3	0.1	-28.3	-29.9	-25.1	**-19.8**
MinT-lasso	-62.1	**-64.4**	-58.9	-51.5	**-0.3**	**0.3**	**0.4**	**0.1**	**0.6**	**0.3**	**0.1**	0.1	**-28.4**	**-30.1**	**-25.2**	**-19.8**
MinTs	**-62.2**	**-64.4**	**-59.0**	**-51.6**	**-0.3**	0.3	0.4	0.1	0.4	0.3	0.1	**0.0**	**-28.5**	-30.1	-25.2	**-19.8**
MinTs-subset	-62.0	-63.8	-58.4	-51.1	0.4	1.1	1.2	0.7	0.5	0.9	0.7	0.5	-28.2	-29.5	-24.6	-19.3
MinTs-intuitive	**-62.2**	**-64.4**	**-59.0**	**-51.6**	**-0.3**	0.3	0.4	0.1	0.4	0.3	0.1	**0.0**	**-28.5**	-30.1	-25.2	**-19.8**
MinTs-lasso	**-62.2**	**-64.4**	-58.9	-51.5	-0.2	0.3	0.4	0.1	**0.2**	**0.2**	0.1	**0.0**	**-28.5**	-30.1	-25.2	**-19.8**
Elasso	-60.9	-63.6	-58.2	-51.1	2.3	2.8	2.3	1.3	3.1	3.1	2.1	1.2	-26.5	-28.3	-23.8	-18.9

5.2 Tourism data

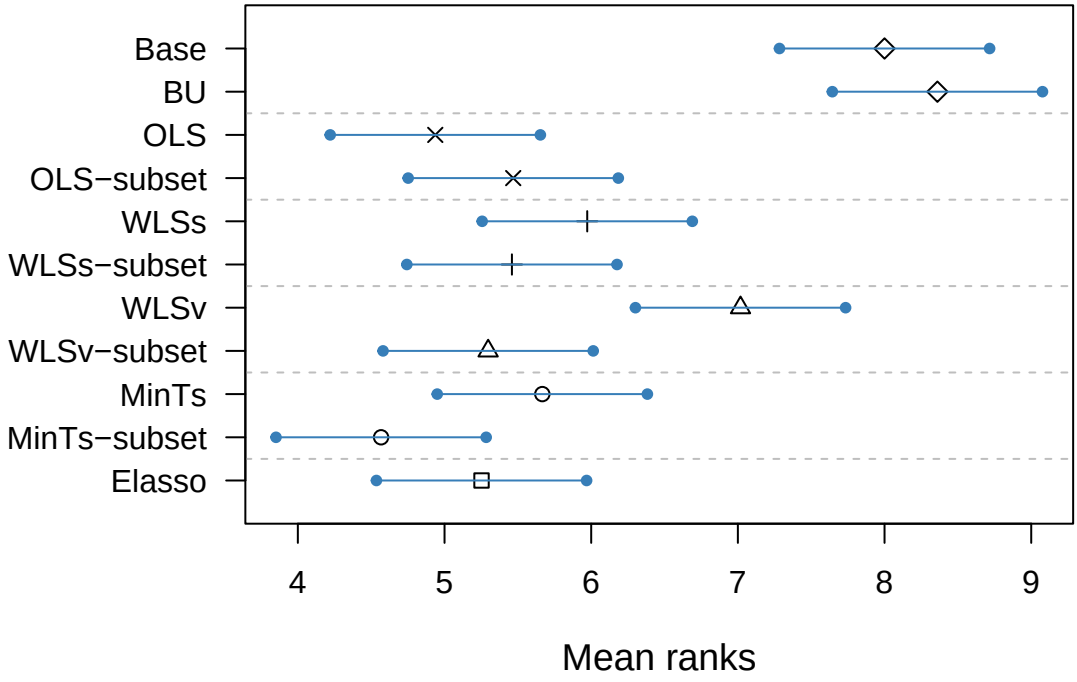


Figure 2: MCB test conducted on the methods examined using the tourism data, the ranks are computed considering all the time series of the hierarchy included in the data set, i.e. a total of 111 series in the hierarchy.

Table 5: Ratio of being retained for each time series after subset selection in 500 simulation data instances.

	Top	A	B	AA	AB	BA	BB	Summary
OLS-subset	0.52	0.54	0.58	0.87	0.89	0.90	0.83	
OLS-intuitive	0.68	0.57	0.61	0.82	0.86	0.84	0.81	
OLS-lasso	0.62	0.52	0.53	1.00	1.00	1.00	1.00	
WLSs-subset	0.53	0.59	0.64	0.89	0.91	0.87	0.89	
WLSs-intuitive	0.65	0.58	0.61	0.86	0.92	0.87	0.88	
WLSs-lasso	0.60	0.58	0.59	1.00	1.00	1.00	1.00	
WLSv-subset	0.52	0.62	0.64	0.88	0.89	0.87	0.89	
WLSv-intuitive	0.64	0.57	0.55	0.87	0.93	0.87	0.92	
WLSv-lasso	0.60	0.60	0.61	1.00	1.00	1.00	1.00	
MinT-subset	0.55	0.56	0.57	0.91	0.92	0.89	0.90	
MinT-intuitive	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MinT-lasso	0.76	0.81	0.80	0.97	0.97	0.97	0.97	
MinTs-subset	0.47	0.46	0.52	0.91	0.92	0.91	0.90	
MinTs-intuitive	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MinTs-lasso	0.63	0.64	0.67	1.00	1.00	1.00	1.00	
Elasso	0.84	0.67	0.69	1.00	1.00	1.00	1.00	

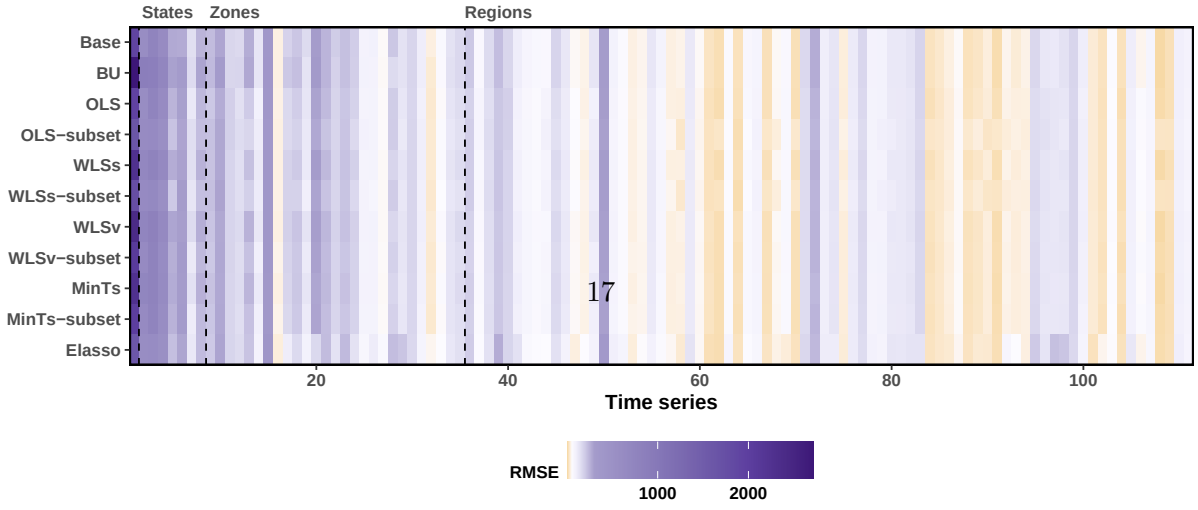


Figure 3: Average reconciliation errors in terms of RMSE (1- to 12-step-ahead) after forecast

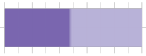
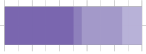
Table 6: Ratio of being retained for each time series after subset selection in 500 simulation data instances (Scenario I).

	Top	A	B	AA	AB	BA	BB	Summary
OLS-subset	0.52	0.79	0.57	0.79	1	0.91	0.85	
OLS-intuitive	0.80	0.90	0.81	0.80	1	0.85	0.86	
OLS-lasso	0.90	1.00	0.68	1.00	1	1.00	1.00	
WLSs-subset	0.85	0.91	0.86	0.90	1	0.97	0.97	
WLSs-intuitive	0.92	0.95	0.67	0.92	1	0.92	0.95	
WLSs-lasso	0.72	1.00	0.72	1.00	1	1.00	1.00	
WLSv-subset	0.50	0.62	0.42	0.19	1	0.81	0.87	
WLSv-intuitive	0.59	0.55	0.49	0.17	1	0.76	0.86	
WLSv-lasso	0.40	1.00	0.41	0.77	1	1.00	1.00	
MinT-subset	0.66	0.90	0.61	0.72	1	0.91	0.93	
MinT-intuitive	1.00	1.00	1.00	1.00	1	1.00	1.00	
MinT-lasso	0.80	0.96	0.84	0.72	1	0.98	0.97	
MinTs-subset	0.57	0.88	0.52	0.67	1	0.89	0.92	
MinTs-intuitive	1.00	1.00	1.00	1.00	1	1.00	1.00	
MinTs-lasso	0.68	1.00	0.66	0.74	1	1.00	1.00	
Elasso	0.82	0.63	0.69	1.00	1	1.00	1.00	

6 Further issues

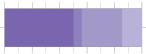
- Update all the results after using a new way to decide λ sequence.

Table 7: Ratio of being retained for each time series after subset selection in 500 simulation data instances (Scenario II).

	Top	A	B	AA	AB	BA	BB	Summary
OLS-subset	0.55	0.04	0.41	0.74	0.78	0.79	0.83	
OLS-intuitive	0.61	0.04	0.52	0.75	0.69	0.69	0.83	
OLS-lasso	0.04	0.35	0.02	1.00	1.00	1.00	1.00	
WLSs-subset	0.45	0.06	0.36	0.81	0.84	0.81	0.87	
WLSs-intuitive	0.61	0.06	0.48	0.75	0.71	0.73	0.84	
WLSs-lasso	0.02	0.33	0.02	1.00	1.00	1.00	1.00	
WLSv-subset	0.54	0.29	0.46	0.91	0.94	0.86	0.89	
WLSv-intuitive	0.59	0.32	0.53	0.82	0.86	0.77	0.86	
WLSv-lasso	0.27	0.42	0.26	1.00	1.00	1.00	1.00	
MinT-subset	0.69	0.64	0.66	0.95	0.96	0.90	0.90	
MinT-intuitive	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MinT-lasso	0.82	0.74	0.83	1.00	0.99	0.97	0.97	
MinTs-subset	0.62	0.63	0.58	0.95	0.96	0.90	0.86	
MinTs-intuitive	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MinTs-lasso	0.68	0.75	0.68	1.00	1.00	1.00	1.00	
Elasso	0.78	0.95	0.68	1.00	1.00	1.00	1.00	

- Simulation: using different base models across the hierarchy.
- Result presentation: visualization aspect.
- Grouped time series.
- More on large hierarchy:

Table 8: Ratio of being retained for each time series after subset selection in 500 simulation data instances (Scenario III).

	Top	A	B	AA	AB	BA	BB	Summary
OLS-subset	0.75	0.45	0.44	0.82	0.79	0.83	0.80	
OLS-intuitive	0.47	0.70	0.69	0.86	0.92	0.90	0.89	
OLS-lasso	0.38	0.01	0.01	1.00	1.00	1.00	1.00	
WLSs-subset	0.08	0.42	0.41	0.87	0.85	0.84	0.89	
WLSs-intuitive	0.06	0.55	0.50	0.66	0.87	0.69	0.88	
WLSs-lasso	0.35	0.03	0.03	1.00	1.00	1.00	1.00	
WLSv-subset	0.31	0.67	0.65	0.88	0.90	0.91	0.90	
WLSv-intuitive	0.34	0.63	0.60	0.80	0.89	0.84	0.87	
WLSv-lasso	0.45	0.35	0.36	1.00	1.00	1.00	1.00	
MinT-subset	0.69	0.78	0.80	0.91	0.91	0.91	0.91	
MinT-intuitive	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MinT-lasso	0.75	0.89	0.86	0.97	0.97	0.97	0.97	
MinTs-subset	0.67	0.74	0.76	0.90	0.89	0.88	0.91	
MinTs-intuitive	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
MinTs-lasso	0.77	0.72	0.73	1.00	1.00	1.00	1.00	
Elasso	0.95	0.64	0.64	1.00	1.00	1.00	1.00	

– Sub hierarchy + Voting

Table 9: Out-of-sample forecast performance (average RMSE) for the tourism data.

	Top				State				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	1158.2	716.6	1279.5	1907.6	452.7	323.3	349.9	424.8	165.5	163.6	160.7	179.7	100.8	89.4	88.2	94.1	148.3	127.9	133.1	152.1
BU	89.1	132.8	53.4	42.0	-4.6	10.3	17.0	19.7	1.1	-2.4	0.4	1.0	0.0	0.0	0.0	0.0	5.7	7.6	7.6	8.5
OLS	-4.7	-0.4	0.5	1.4	-3.0	-3.9	-1.6	-1.5	-2.1	-4.2	-5.6	-7.5	1.0	-0.4	-1.9	-3.2	-1.0	-2.1	-2.7	-3.6
OLS-subset	-4.7	8.0	**-1.4**	**-14.1**	-3.0	5.5	0.3	**-7.9**	-2.1	-1.5	-3.7	**-8.7**	1.0	1.7	-0.1	-2.3	-1.0	1.7	-1.2	**-6.5**
OLS-intuitive	-4.7	-0.4	0.5	1.4	-3.0	-3.9	-1.6	-1.5	-2.1	-4.2	-5.6	-7.5	1.0	-0.4	-1.9	-3.2	-1.0	-2.1	-2.7	-3.6
OLS-lasso	-4.7	-0.4	0.5	1.4	-3.0	-3.9	-1.6	-1.5	-2.1	-4.2	-5.6	-7.5	1.0	-0.4	-1.9	-3.2	-1.0	-2.1	-2.7	-3.6
WLSs	25.1	55.2	20.8	19.1	-15.8	-5.0	3.5	6.2	-5.9	-5.4	-4.7	-5.0	-0.2	-0.8	-1.6	-2.2	-3.0	-0.1	0.3	0.9
WLSs-subset	25.1	**18.7**	**0.8**	**-7.8**	-15.8	-2.7	**-2.1**	**-6.2**	-5.9	-4.1	**-4.8**	**-8.5**	-0.2	0.3	-1.0	**-2.5**	-3.0	**-0.6**	**-2.1**	**-5.5**
WLSs-intuitive	25.1	55.2	20.8	19.1	-15.8	-5.0	3.5	6.2	-5.9	-5.4	-4.7	-5.0	-0.2	-0.8	-1.6	-2.2	-3.0	-0.1	0.3	0.9
WLSs-lasso	25.1	55.2	20.8	19.1	-15.8	-5.0	3.5	6.2	-5.9	-5.4	-4.7	-5.0	-0.2	-0.8	-1.6	-2.2	-3.0	-0.1	0.3	0.9
WLSv	38.2	76.2	29.6	25.6	-17.4	-3.1	7.0	9.9	-5.0	-4.3	-3.1	-3.2	-4.2	-1.6	-1.8	-2.1	-3.9	1.3	2.0	2.8
WLSv-subset	38.2	**34.5**	**10.7**	**8.5**	-17.4	**-8.8**	**-0.8**	**1.4**	-5.0	**-5.5**	**-5.3**	**-6.7**	-4.1	**-2.0**	**-2.6**	**-3.4**	-3.9	**-2.3**	**-2.0**	**-2.2**
WLSv-intuitive	38.2	76.2	29.6	25.6	-17.4	-3.1	7.0	9.9	-5.0	-4.3	-3.1	-3.2	-4.2	-1.6	-1.8	-2.1	-3.9	1.3	2.0	2.8
WLSv-lasso	38.2	76.2	29.6	25.6	-17.4	-3.1	7.0	9.9	-5.0	-4.3	-3.1	-3.2	-4.2	-1.6	-1.8	-2.1	-3.9	1.3	2.0	2.8
MinTs	20.6	53.6	21.6	19.0	**-22.2**	-7.2	3.5	6.3	**-12.1**	-6.6	-5.1	-5.3	**-5.3**	-2.6	-2.8	-3.1	-8.6	-1.8	-0.3	0.4
MinTs-subset	20.6	**20.0**	**6.4**	**5.6**	**-22.2**	**-11.3**	**-2.5**	**-0.1**	**-12.1**	**-7.5**	**-6.4**	**-7.8**	**-5.3**	**-2.9**	**-3.2**	**-3.9**	-8.6	**-4.5**	**-3.2**	**-3.3**
MinTs-intuitive	20.6	53.6	21.6	19.0	**-22.2**	-7.2	3.5	6.3	**-12.1**	-6.6	-5.1	-5.3	**-5.3**	-2.6	-2.8	-3.1	-8.6	-1.8	-0.3	0.4
MinTs-lasso	20.6	53.6	21.6	19.0	**-22.2**	-7.2	3.5	6.3	**-12.1**	-6.6	-5.1	-5.3	**-5.3**	-2.6	-2.8	-3.1	-8.6	-1.8	-0.3	0.4
Elasso	**-84.5**	**-50.4**	**-16.3**	**-16.4**	-18.3	0.6	**-9.0**	**-11.4**	-7.8	**-8.8**	**-7.5**	**-10.4**	2.9	1.6	4.1	0.3	**-10.2**	-4.4	**-3.2**	**-6.7**

Table 10: Number of time series retained after subset selection and the optimal hyperparameter values for the tourism data.

	Number of time series retained					Optimal parameters		
	Top	State	Zone	Region	Total	λ_0	λ_1	λ_2
None	1	7	27	76	111	0.00	0.00	0.00
OLS-subset	1	2	13	76	92	27.98	-	10.00
WLSs-subset	1	1	15	76	93	18.73	-	10.00
WLSv-subset	1	7	27	76	111	0.03	-	0.01
MinTs-subset	1	7	27	76	111	0.05	-	0.01
Elasso	1	4	0	8	13	-	71759.21	-

7 References

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.

Yang, Yi, and Hui Zou. 2014. "A Fast Unified Algorithm for Solving Group-Lasso Penalize Learning Problems." *Statistics and Computing* 25 (6): 1129–41. https://doi.org/10.1007/s11222-014-9498-5.