Results of Forecast Reconciliation with Subset Selection

Xiaoqian Wang

2023-05-31

1 Forecast Reconciliation with Subset Selection

With two unbiasedness conditions, the trace minimization (MinT) problem can be reformulated in terms of a **linear equality constrained least squares problem** as follow:

$$\min_{\boldsymbol{G}} \quad \frac{1}{2}\left(\hat{\boldsymbol{y}}_h-\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h\right)'\boldsymbol{W}_h^{-1}\left(\hat{\boldsymbol{y}}_h-\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}_h\right) \quad \text{s.t. } \boldsymbol{G}\boldsymbol{S}=\boldsymbol{I}_{n_b}.$$

If we consider the optimization problem with ℓ_0, ℓ_1, and ℓ_2 penalizations, we have:

$$\begin{aligned}
\min_{\boldsymbol{G}} \quad & \frac{1}{2}\left(\hat{\boldsymbol{y}}_h-\left(\hat{\boldsymbol{y}}_h'\otimes\boldsymbol{S}\right)\operatorname{vec}(\boldsymbol{G})\right)'\boldsymbol{W}_h^{-1}\left(\hat{\boldsymbol{y}}_h-\left(\hat{\boldsymbol{y}}_h'\otimes\boldsymbol{S}\right)\operatorname{vec}(\boldsymbol{G})\right) \\
& \qquad +\lambda_0\sum_{j=1}^{n}\|\boldsymbol{G}_{\cdot j}\|_0+\lambda_1\left\|\operatorname{vec}\left(\boldsymbol{G}-\boldsymbol{G}^0\right)\right\|_1+\lambda_2\left\|\operatorname{vec}\left(\boldsymbol{G}-\boldsymbol{G}^0\right)\right\|_2^2 \\
\text{s.t.} \quad & \boldsymbol{G}\boldsymbol{S}=\boldsymbol{I}_{n_b},
\end{aligned}$$

where $\boldsymbol{G}^0$ can be a benchmark $\boldsymbol{G}$ matrix estimated by MinT or other methods, such as bottom-up and top-down. The optimization problem can be formulated to a Big-M based MIP formulation, which is a Mixed Integer Quadratic Program (MIQP).

1.1 Small hierarchy

Estimate the whole $\boldsymbol{G}$ matrix.

1.1.1 Data simulation

Structure:

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

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

$$\boldsymbol{b}_t = \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \boldsymbol{\eta}_t$$

where $\boldsymbol{\mu}_t, \boldsymbol{\gamma}_t$, and $\boldsymbol{\eta}_t$ are the trend, seasonal, and error components, respectively,

$$\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{v}_t + \boldsymbol{\varrho}_t, & \boldsymbol{\varrho}_t &\sim \mathcal{N}\left(\mathbf{0}, \sigma_\varrho^2 \boldsymbol{I}_4\right), \\
\boldsymbol{v}_t &= \boldsymbol{v}_{t-1} + \boldsymbol{\zeta}_t, & \boldsymbol{\zeta}_t &\sim \mathcal{N}\left(\mathbf{0}, \sigma_\zeta^2 \boldsymbol{I}_4\right), \\
\boldsymbol{\gamma}_t &= -\sum_{i=1}^{s-1} \boldsymbol{\gamma}_{t-i} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim \mathcal{N}\left(\mathbf{0}, \sigma_\omega^2 \mathbf{I}_4\right),
\end{aligned}$$

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

- $\sigma_\varrho^2 = 2, \sigma_\zeta^2 = 0.007$, and $\sigma_\omega^2 = 7$.
- $s = 4$ for quarterly data, $n = 180$, $h = 16$.
- The initial values for $\boldsymbol{\mu}_0, \boldsymbol{v}_0, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ were generated independently from a multivariate normal distribution with mean zero and covariance matrix, $\Sigma_0 = I_4$.
- Each component of $\boldsymbol{\eta}_t$ 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.

1.1.2 Model specification

1. **Hyperparameter:**

- Set $\lambda_1 = \lambda_2 = 0$
- λ_0 is selected by minimizing sum of squared residuals in the training set.
  - Set $\lambda_{\max}$ to $\frac{1}{n_b}\frac{1}{2}\hat{\boldsymbol{y}}_h' \boldsymbol{W}_h^{-1} \hat{\boldsymbol{y}}_h$.
  - Similarly to the `glmnet` package, we select a minimum value $\lambda_{\min} = \epsilon\lambda_{\max}$, and construct a sequence of $K-1$ values of λ decreasing from $\lambda_{\max}$ to $\lambda_{\min}$ **on the log scale**. $\epsilon = 10^{-4}$ and $K = 20$.
  - Select the best value of λ_0 from $0, \lambda_{\min}, \ldots, \lambda_{\max}$ 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. (Avoid cross-validation)

2. **Four scenarios:**

- S0: ETS
  - ETS models are used to generate base forecasts.
- S1: D-AA
  - Base forecasts (and also fitted values) of **series AA** multiplied by 1.5 to achieve deterioration.
- S2: D-A
  - Base forecasts (and also fitted values) of **series A** multiplied by 1.5 to achieve deterioration.
- S3: D-Total
  - Base forecasts (and also fitted values) of **series Total** multiplied by 1.5 to achieve deterioration.

1.1.3 Results

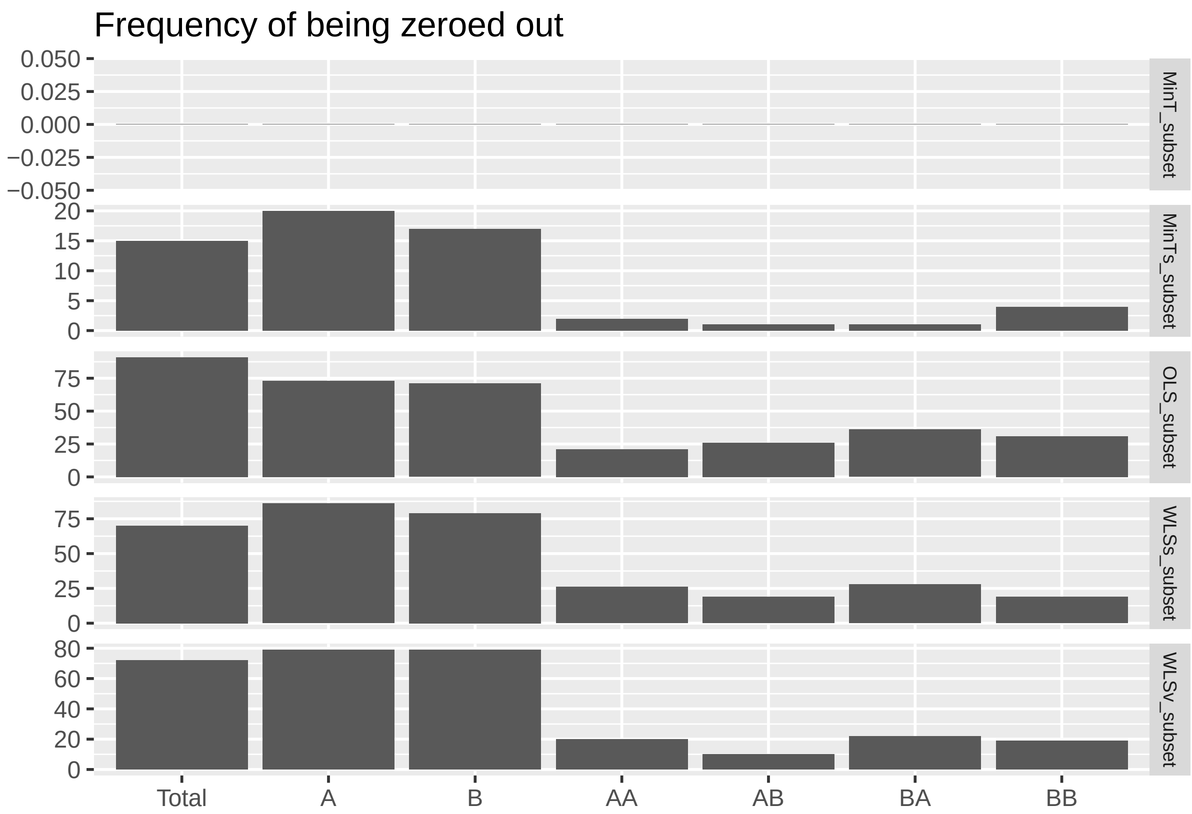
1.1.3.1 S0: ETS

1. RMSE results

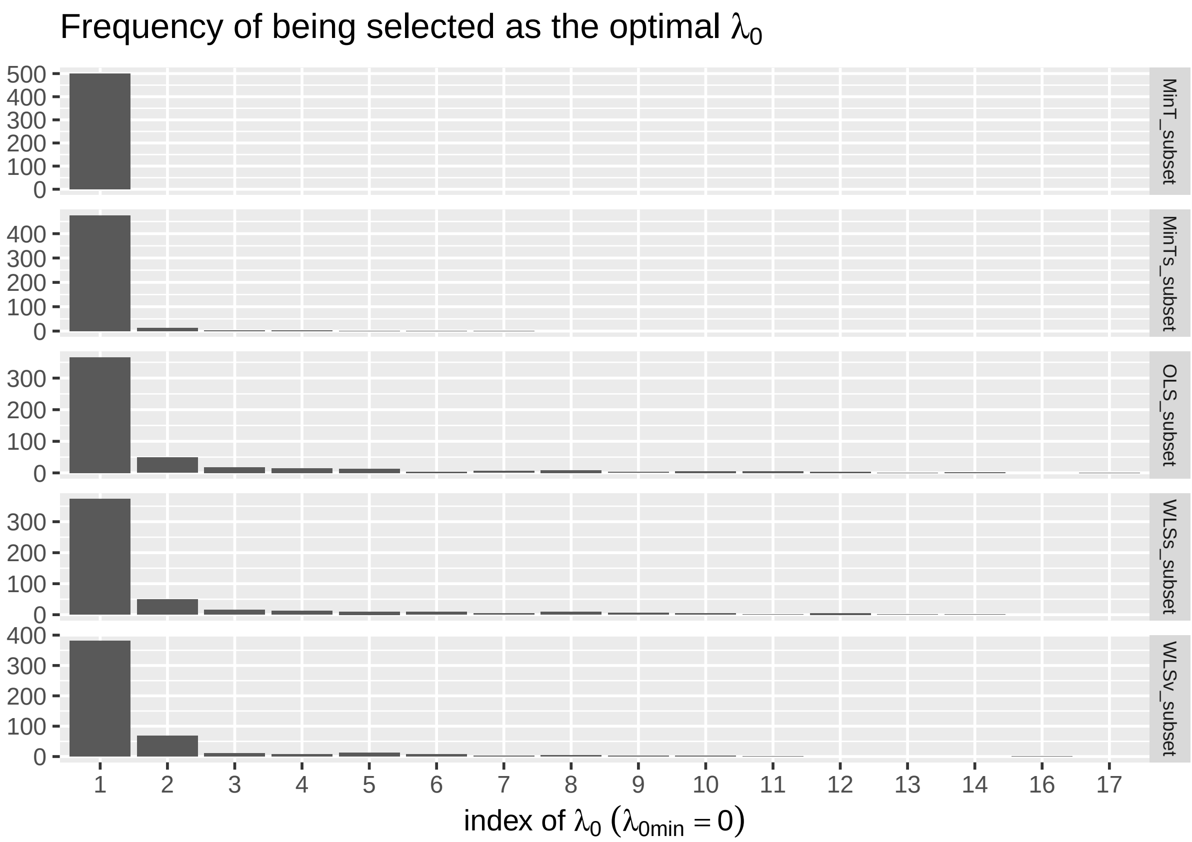
	Top				Middle				Bottom				Average			
Method	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	9.61	10.73	12.59	**15.58**	6.33	7.26	8.61	10.83	**4.20**	4.92	5.93	7.52	5.58	6.41	7.65	9.62
BU	**9.51**	10.78	12.67	15.68	6.32	7.25	8.62	10.83	**4.20**	4.92	5.93	7.52	**5.56**	6.42	7.66	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	9.55	10.72	12.60	15.60	6.33	7.23	8.59	10.81	4.21	4.91	5.92	7.51	5.58	6.40	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-subset	9.53	10.73	12.62	15.62	6.32	7.24	8.60	10.81	4.20	4.91	5.92	7.51	5.57	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-subset	9.53	10.73	12.62	15.62	6.31	7.24	8.60	10.81	4.20	4.91	5.92	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

	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
OLS-lasso	11.41	32.13	33.24	34.84	8.91	26.96	27.49	28.45	8.06	21.96	22.23	22.80	8.78	24.84	25.31	26.13
WLSs-lasso	11.69	32.43	33.53	35.11	8.78	27.10	27.63	28.59	7.24	22.11	22.38	22.95	8.32	25.01	25.47	26.30
WLSv-lasso	11.59	32.43	33.54	35.12	8.68	27.11	27.64	28.60	7.05	22.15	22.42	22.99	8.17	25.04	25.50	26.33
MinT-lasso	11.74	32.19	33.31	34.90	8.51	27.20	27.72	28.69	5.95	22.58	22.84	23.42	7.51	25.27	25.73	26.56
MinTs-lasso	11.68	32.18	33.31	34.90	8.50	27.19	27.71	28.68	5.94	22.56	22.83	23.41	7.49	25.26	25.72	26.55
Emp-lasso	**9.74**	**11.00**	**12.87**	**15.81**	**6.47**	**7.45**	**8.79**	**10.96**	**4.37**	**5.11**	**6.09**	**7.64**	**5.73**	**6.62**	**7.83**	**9.76**

2. Selection results



3. Hyperparameters results



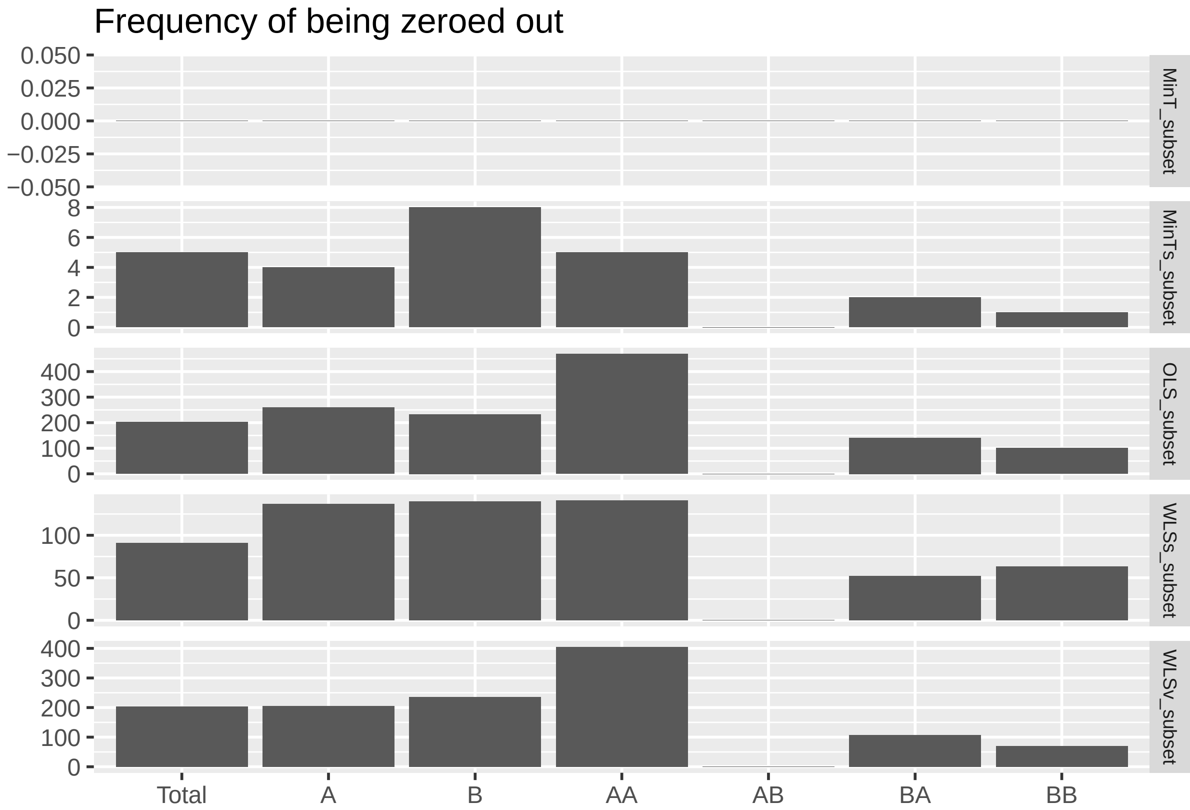
1.1.3.2 S1: D-AA

1. RMSE results

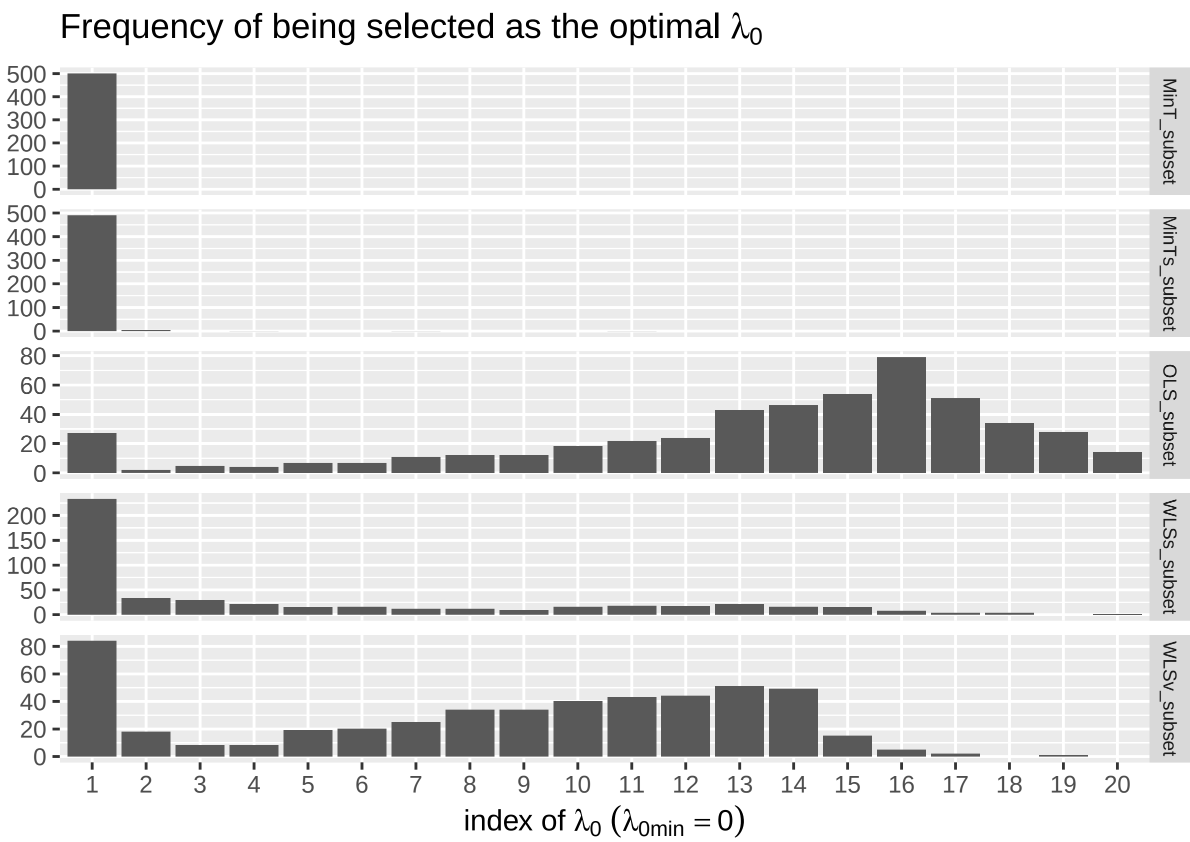
Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	9.61	**10.73**	**12.59**	**15.58**	**6.33**	**7.26**	**8.61**	**10.83**	6.38	7.47	8.34	9.75	6.83	7.88	9.02	10.89
BU	15.16	18.08	19.35	21.64	10.02	11.74	12.75	14.55	6.38	7.47	8.34	9.75	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	9.56	10.74	12.63	15.63	6.40	7.29	8.64	10.85	4.30	5.03	6.04	7.61	5.65	6.49	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	10.27	11.53	13.36	16.29	7.29	8.11	9.38	11.49	5.88	6.44	7.32	8.76	6.91	7.64	8.77	10.61
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	9.56	10.78	12.67	15.66	6.40	7.33	8.68	10.88	4.33	5.05	6.05	7.62	5.67	6.52	7.74	9.70
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	12.69	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

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
OLS-lasso	11.59	33.29	34.37	35.92	9.13	27.48	27.99	28.92	8.33	21.88	22.17	22.74	9.03	25.11	25.57	26.39
WLSs-lasso	12.00	33.43	34.53	36.07	9.16	27.60	28.11	29.04	7.57	22.00	22.28	22.86	8.66	25.23	25.69	26.51
WLSv-lasso	11.53	33.26	34.35	35.91	8.93	27.56	28.06	29.00	7.39	22.14	22.42	23.00	8.42	25.28	25.74	26.56
MinT-lasso	11.82	33.17	34.27	35.84	8.71	27.70	28.20	29.15	6.16	22.35	22.62	23.19	7.70	25.42	25.88	26.70
MinTs-lasso	11.78	33.17	34.27	35.84	8.68	27.68	28.18	29.13	6.18	22.34	22.61	23.19	7.70	25.41	25.87	26.69
Emp-lasso	**9.74**	**11.00**	**12.87**	**15.81**	**6.47**	**7.45**	**8.79**	**10.96**	**4.36**	**5.10**	**6.09**	**7.64**	**5.73**	**6.61**	**7.83**	**9.75**

2. Selection results



3. Hyperparameters results



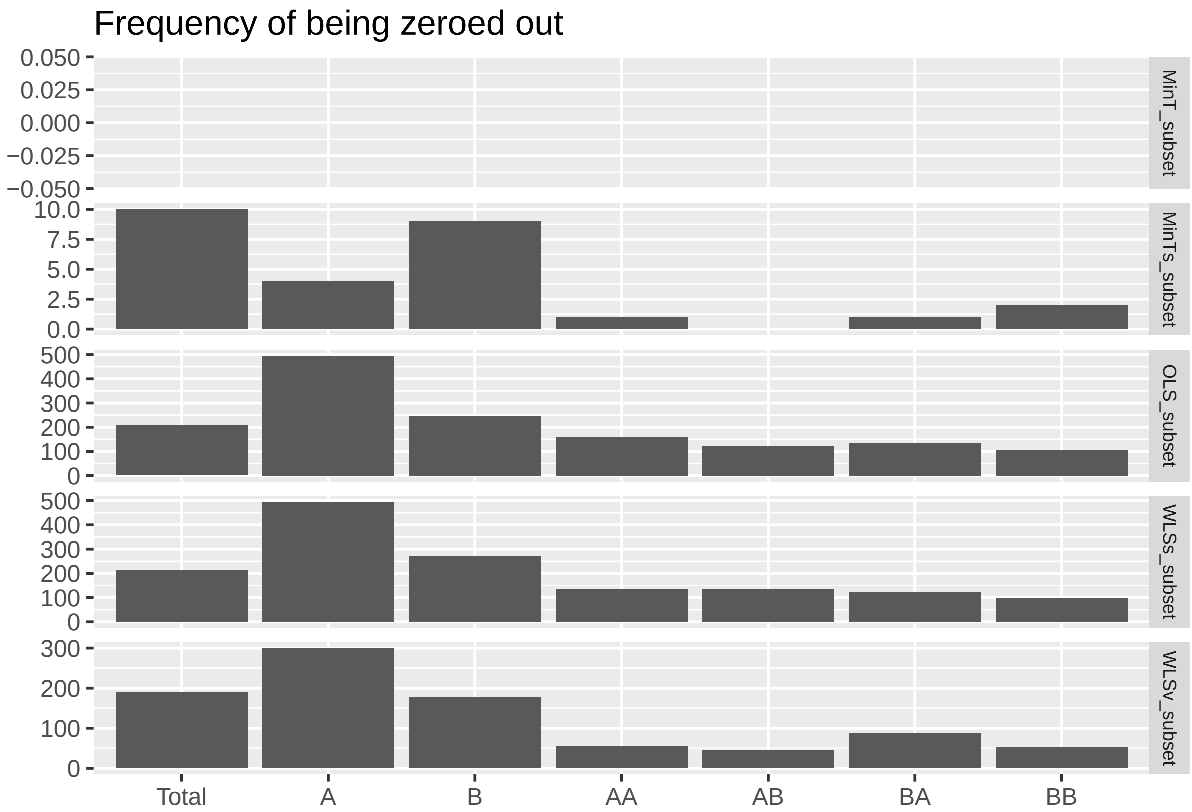
1.1.3.3 S2: D-A

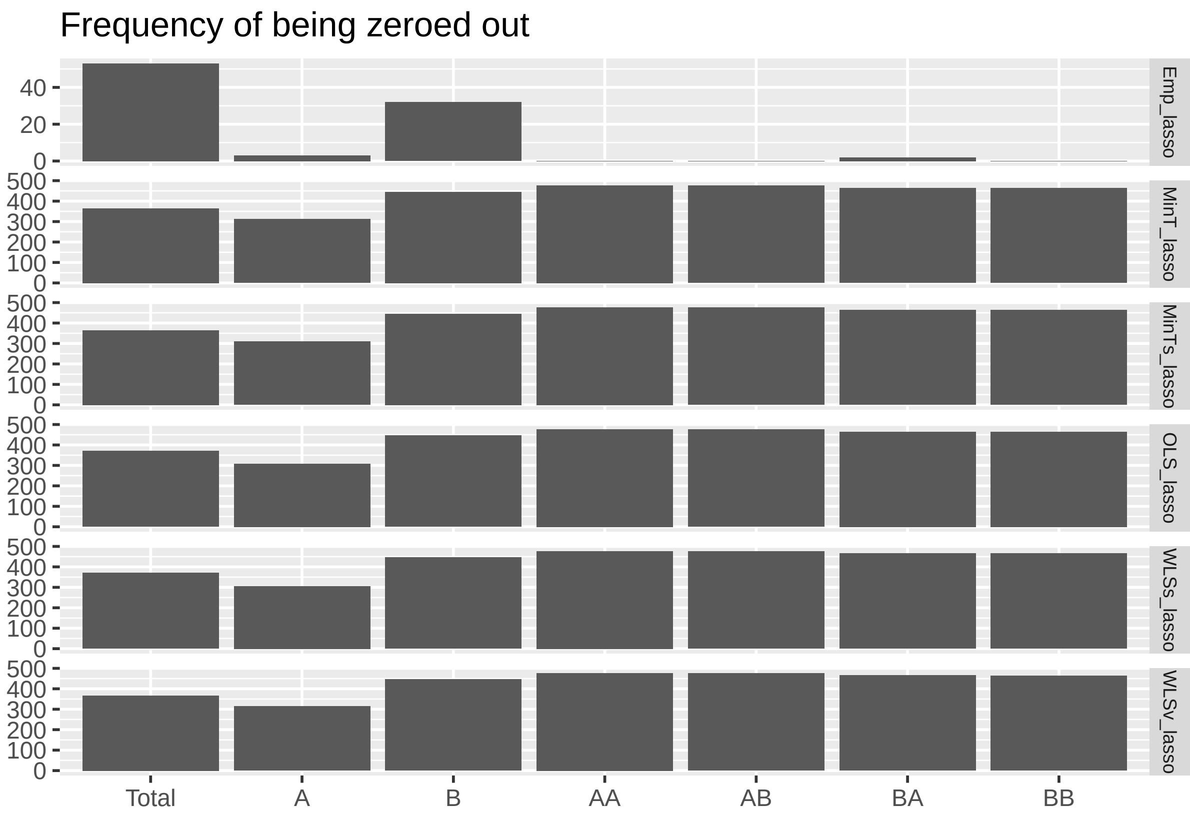
1. RMSE results

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	9.54	10.75	12.64	15.64	6.38	7.31	8.65	10.85	4.24	4.96	5.97	7.55	5.61	6.46	7.69	9.65
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	9.57	10.77	12.65	15.65	6.38	7.31	8.66	10.86	4.25	4.97	5.98	7.56	5.62	6.47	7.70	9.66
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	9.59	10.84	12.71	15.70	6.41	7.37	8.72	10.91	4.24	4.96	5.97	7.56	5.62	6.49	7.72	9.68
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	7.67	**9.63**

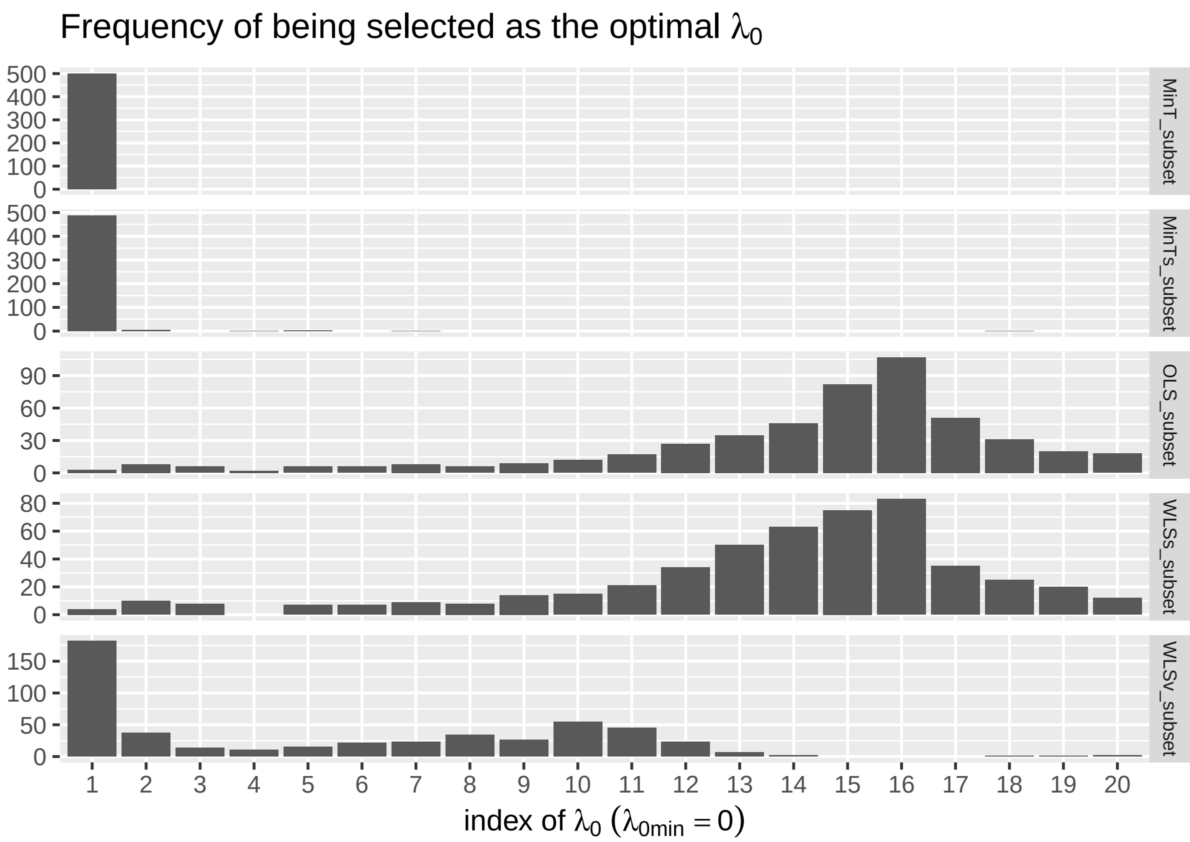
Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
OLS-lasso	10.59	33.95	34.98	36.51	9.01	25.84	26.44	27.46	8.83	21.99	22.27	22.82	9.13	24.80	25.28	26.10
WLSs-lasso	11.17	34.09	35.12	36.61	8.86	25.93	26.53	27.54	7.84	22.09	22.37	22.92	8.61	24.90	25.38	26.20
WLSv-lasso	11.21	34.52	35.51	36.97	9.10	26.36	26.93	27.92	7.18	22.32	22.59	23.15	8.30	25.22	25.68	26.48
MinT-lasso	11.90	34.40	35.42	36.87	9.15	26.38	26.94	27.93	6.32	22.68	22.94	23.50	7.92	25.41	25.86	26.67
MinTs-lasso	11.83	34.40	35.42	36.89	9.11	26.37	26.93	27.92	6.30	22.67	22.93	23.49	7.89	25.40	25.86	26.67
Emp-lasso	**9.74**	**11.00**	**12.87**	**15.81**	**6.46**	**7.44**	**8.79**	**10.96**	**4.38**	**5.12**	**6.10**	**7.65**	**5.74**	**6.62**	**7.84**	**9.76**

2. Selection results





3. Hyperparameters results



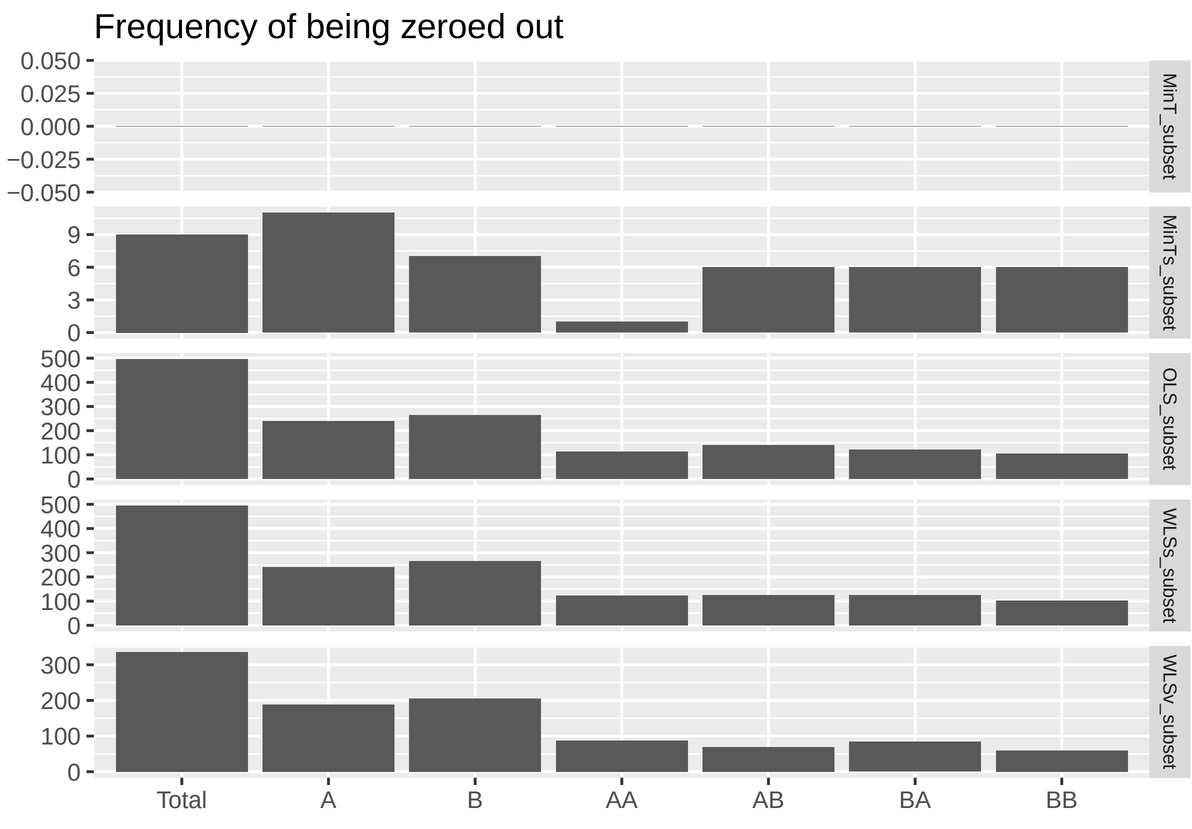
1.1.3.4 S3: D-Total

1. RMSE results

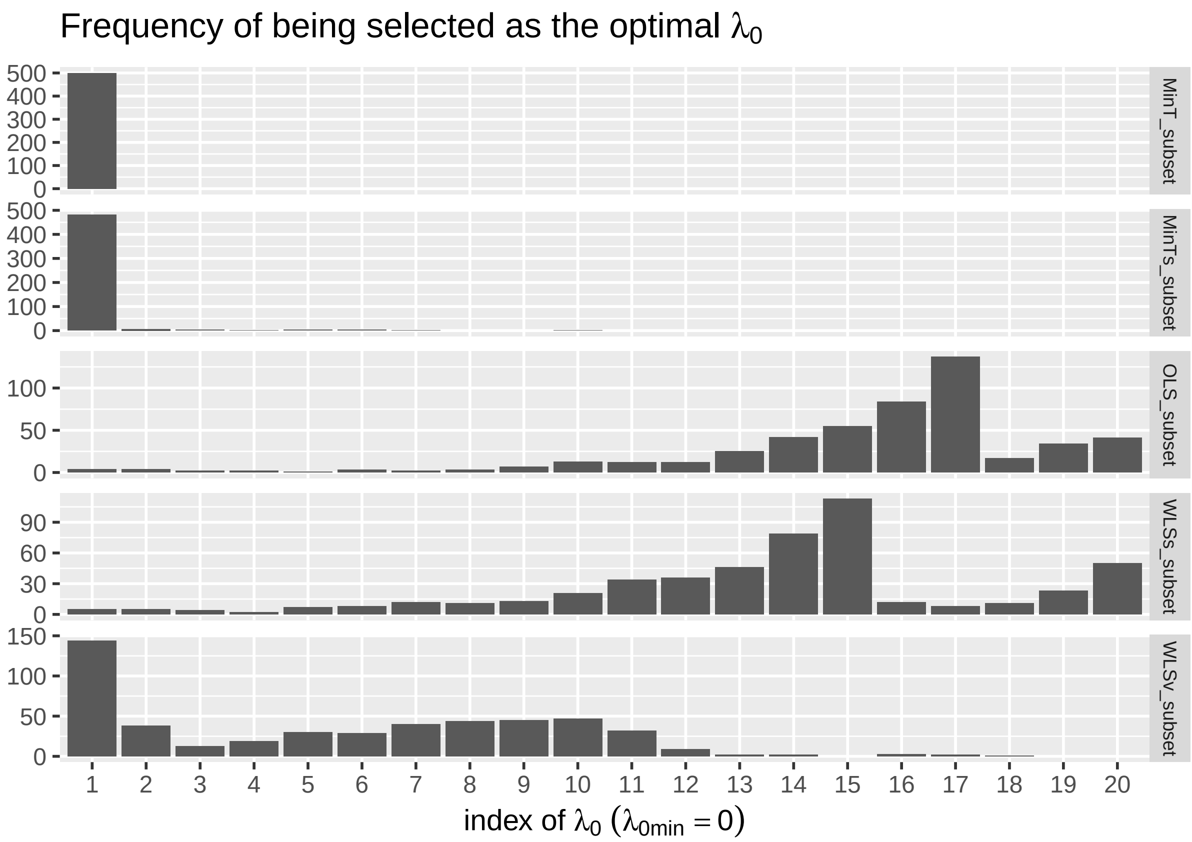
Method	Top				Middle				Bottom				Average			
	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16	h=1	1-4	1-8	1-16
Base	25.01	30.26	30.88	32.34	6.33	7.26	**8.61**	10.83	**4.20**	**4.92**	**5.93**	**7.52**	7.78	9.20	10.26	12.01
BU	9.51	**10.78**	12.67	15.68	6.32	**7.25**	8.62	**10.83**	**4.20**	**4.92**	**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	9.49	10.81	12.72	15.76	**6.31**	7.28	8.64	10.86	4.21	4.93	5.94	7.54	5.56	6.44	7.68	9.66
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	9.50	10.80	12.71	15.75	6.32	7.28	8.64	10.86	4.21	4.93	5.94	7.54	5.57	6.44	7.68	9.66
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	9.59	10.86	12.74	15.74	6.35	7.30	8.65	10.86	4.21	4.93	5.94	7.53	5.59	6.45	7.69	9.66
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	**9.45**	10.78	**12.66**	**15.65**	6.31	7.28	8.64	10.83	4.21	4.93	5.94	7.52	**5.56**	6.44	7.67	**9.63**

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
OLS-lasso	11.00	26.32	27.68	29.65	10.89	27.42	27.90	28.83	10.46	22.42	22.69	23.22	10.66	24.40	24.89	25.74
WLSs-lasso	11.84	27.71	28.98	30.85	10.01	27.83	28.32	29.24	8.77	22.57	22.83	23.37	9.56	24.81	25.28	26.11
WLSv-lasso	13.01	29.21	30.40	32.14	9.70	28.34	28.82	29.73	7.92	22.86	23.12	23.66	9.16	25.33	25.79	26.60
MinT-lasso	13.95	30.29	31.44	33.11	9.91	28.68	29.15	30.05	6.97	23.39	23.64	24.17	8.81	25.89	26.33	27.13
MinTs-lasso	13.86	30.26	31.40	33.08	9.89	28.68	29.15	30.05	6.98	23.36	23.61	24.15	8.79	25.87	26.31	27.11
Emp-lasso	**9.74**	**10.99**	**12.87**	**15.80**	**6.47**	**7.45**	**8.79**	**10.96**	**4.42**	**5.17**	**6.15**	**7.69**	**5.76**	**6.66**	**7.86**	**9.78**

2. Selection results



3. Hyperparameters results



1.2 Large hierarchy

1.2.1 Strategy 1: direct

Idea: Directly estimate the whole $\boldsymbol{G}$ matrix.

- MIP solvers: Lack of scalability. The best subset selection is an NP-hard problem, which is computationally intensive.
- L0Group: $\boldsymbol{W}_h^{-1}$ and unbiasedness constraints.
- L0Group modification
  - Branch-and-Bound algorithm

1.2.2 Strategy 2: indirect

Idea: Let $\boldsymbol{G} = \boldsymbol{G}^{*}\boldsymbol{A}$, where $\boldsymbol{G}^{*}$ is MinT solution, and $\boldsymbol{A}$ is an $n \times n$ diagonal matrix. Then the aim is to estimate diagonal elements of $\boldsymbol{A}$.

Setup: λ_0 and λ_2 $(10^{-4}, 10^{-2}, 10^{0}, 10^{2}, 10^{4})$ are selected by minimizing sum of squared residuals, $\lambda_1 = 0$

1. Under unbiasedness constraints.

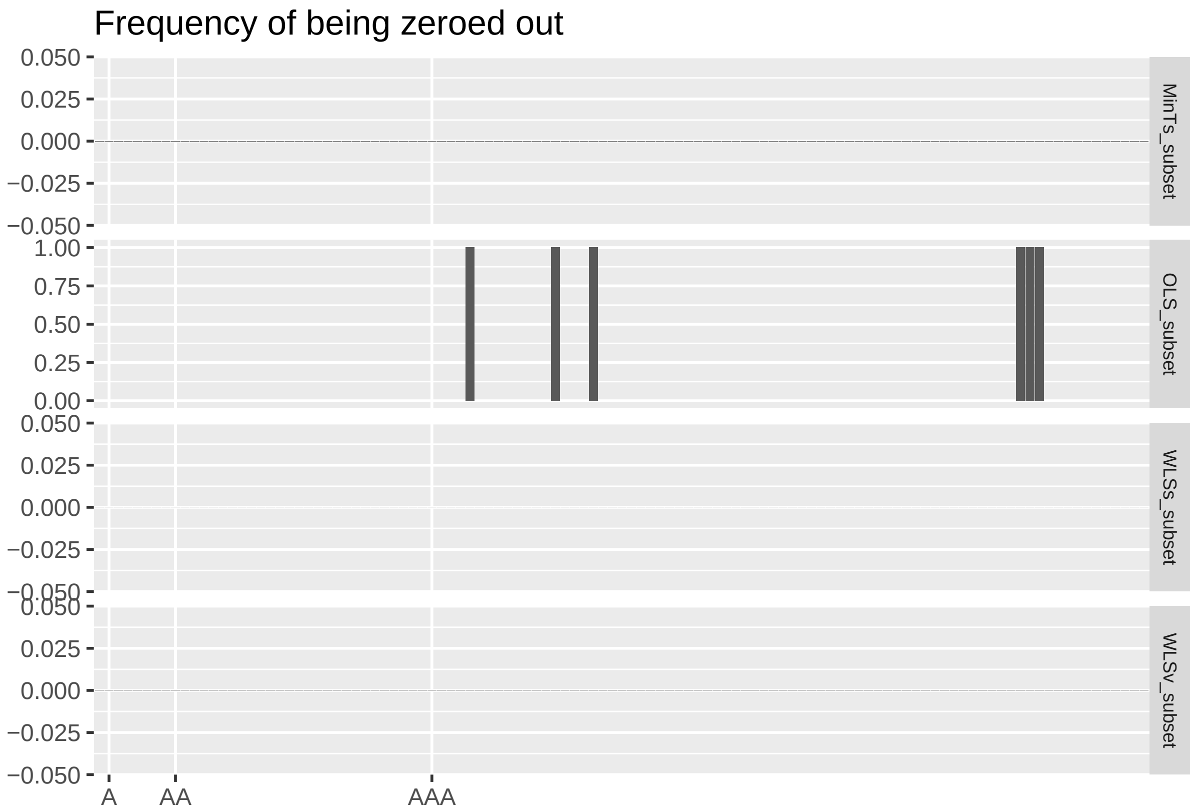
- Small hierarchy
  - **Results:** For simulation data (S1, S2, and S3), the method did not play a role in subset selection, thus giving same results as correponding MinT methods.
  - It is extremely difficult to get a solution of $\boldsymbol{A}$ other than $\boldsymbol{I}$.
- Large hierarchy
  - **Results:** For tourism data (Total/State/Zone/Region: 4 levels, 111 series in total), only `OLS_subset` did subset selection (remove 6 bottom-level series: indices 40, 49, 53, 98, 99, 100), but it achieved the same RMSE performance as `OLS` for every horizon.
  - It is difficult to get a solution of $\boldsymbol{A}$ other than $\boldsymbol{I}$. Even when we get an estimate of $\boldsymbol{A}$ with some of the diagonal elements being zero, the initial weights on the "removed" series are assigned to other time series in the hierarchy to obtain updated bottom-level series. This means the "removed" series are represented using the linear combination of other series.

$$\boldsymbol{G}\hat{\boldsymbol{y}} = [\boldsymbol{g}_{\cdot 1}, \boldsymbol{g}_{\cdot 2}, \cdots, \boldsymbol{g}_{\cdot n}]\, \hat{\boldsymbol{y}};$$

$$\boldsymbol{G}\boldsymbol{A}\hat{\boldsymbol{y}} = [\boldsymbol{g}_{\cdot 1}, \boldsymbol{g}_{\cdot 2}, \cdots, \boldsymbol{g}_{\cdot n}] \begin{bmatrix} a_1\hat{y}_1 \\ a_2\hat{y}_2 \\ \cdots \\ a_n\hat{y}_n \end{bmatrix}.$$

Method	Top				State				Zone				Region				Average			
	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	1448.95	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
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	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
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

Method	Top				State				Zone				Region				Average			
	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
OLS-lasso	1309.50	803.75	1331.02	1970.92	400.12	488.62	705.36	766.14	203.48	**208.74**	279.74	299.72	133.84	131.58	159.62	164.14	178.16	178.92	233.81	251.36
WLSs-lasso	1519.70	930.48	1409.67	2062.30	379.99	458.43	670.68	737.43	156.99	213.76	292.68	312.38	100.90	120.97	154.89	160.63	144.93	172.11	232.24	251.05
WLSv-lasso	1704.90	1045.91	1484.87	2145.77	372.14	449.55	659.40	727.33	158.85	215.85	295.52	315.01	96.82	120.30	154.52	160.36	143.76	172.64	232.64	251.62
MinTs-lasso	1597.19	978.43	1440.52	2096.91	**352.04**	**440.19**	654.11	720.40	**149.05**	213.44	295.19	314.81	**95.88**	**119.93**	154.61	160.49	**138.49**	**170.60**	231.89	250.79
Emp-lasso	**972.50**	**616.85**	**1223.05**	**1833.14**	917.78	689.10	**614.96**	**611.71**	370.83	254.09	**241.57**	**240.90**	171.26	124.20	**124.87**	**123.16**	274.10	195.86	**194.06**	**198.01**

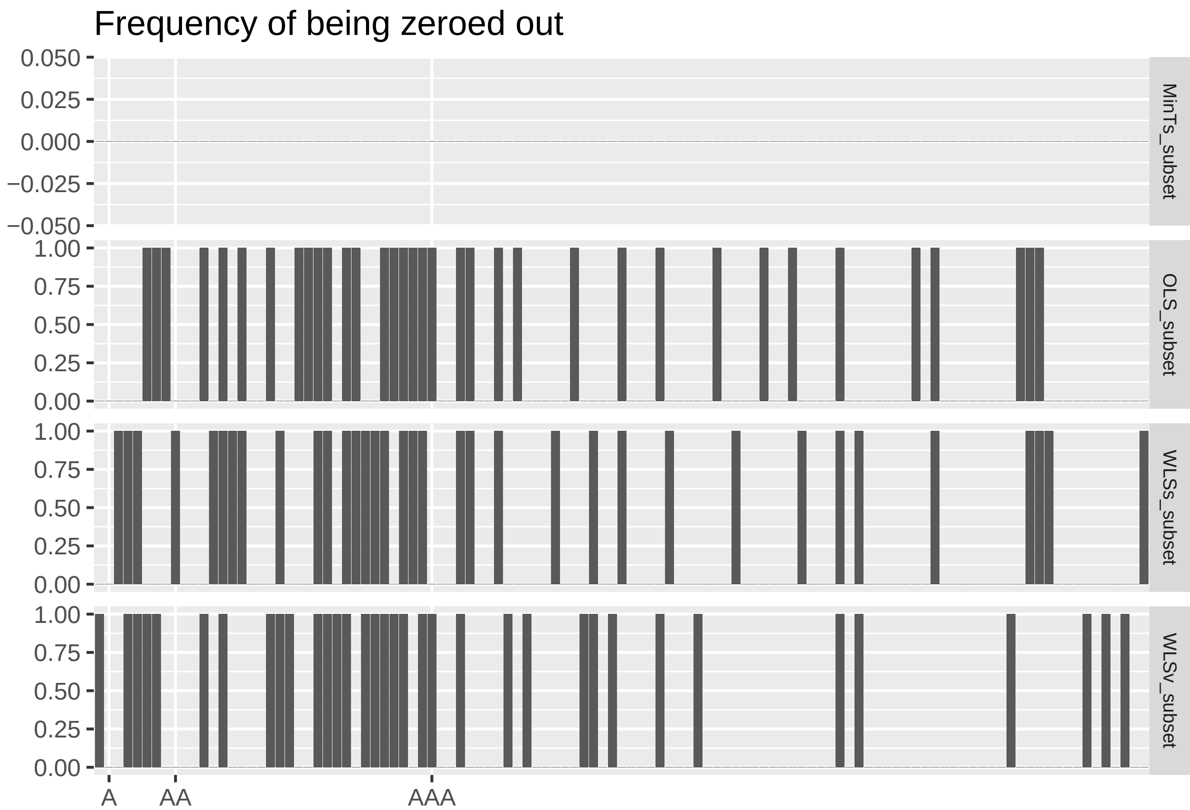


2. Remove unbiasedness constraints.

   Results based on tourism data (Total/State/Zone/Region: 4 levels, 111 series in total)

- `OLS_subset`, `WLSs_subset`, and `WLSv_subset` did subset selection. They give almost the same performance as the corresponding MinT methods for $h = 1$, but extremely worse performance for $h > 1$.

	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	1158.16	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**	832.57	1568.89	2233.61	438.89	453.24	837.32	863.36	162.12	241.85	354.25	372.58	101.79	191.86	273.84	289.59	146.75	226.27	340.60	363.47
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	1448.95	2095.59	3539.62	4246.34	381.21	479.40	737.56	808.70	155.70	205.17	292.73	306.46	100.60	119.07	149.95	152.44	143.85	180.54	252.28	268.17
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	1600.65	1333.64	1429.02	2073.60	374.05	418.79	590.88	644.37	157.23	228.36	298.70	311.99	96.58	132.14	166.29	174.47	142.38	184.45	236.65	254.66
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



- Naturally, when implementing subset selection on forecasts for each horizon, they give almost the same performance as the corresponding MinT methods for all horizons.

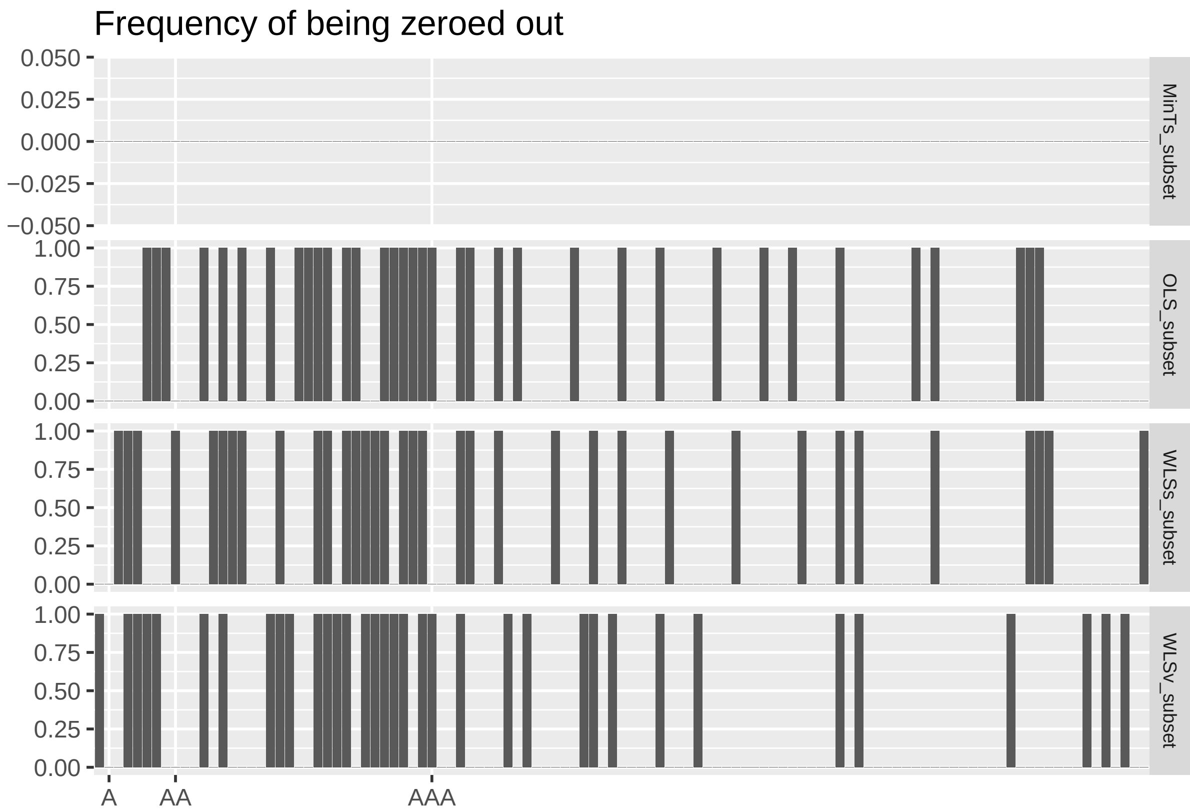
1.2.3 Strategy 3: two-step

Idea: Implement selection first, and then estimation. (MIP solvers are not scalable, but QP solvers are scalable.)

- Step 1: obtain output of binary variables $\boldsymbol{z}$ using **Strategy 2 without unbiasedness constraints**.
- Step 2: directly use $\hat{\boldsymbol{z}}$ to solve the **QP problem with unbiasedness constraints**.

Results: Sometimes cannot find the estimated $\boldsymbol{G}$ in step 2 using the z of step 1. Even if step 2 can find the estimated $\boldsymbol{G}$, it performs worse than the corresponding MinT methods, but its forecasts for large horizons are still in a reasonable range.

	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	1158.16	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	1158.16	716.62	**1279.50**	1907.61	650.81	413.35	415.88	478.21	204.59	186.03	183.47	198.35	115.63	100.90	98.68	104.23	180.41	146.86	149.95	166.96
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	1158.16	716.62	1279.50	1907.61	496.80	352.46	383.51	442.86	170.77	186.86	187.77	211.08	102.03	98.30	97.91	104.90	153.16	141.44	148.42	168.28
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	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
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



1.2.4 Strategy 3: $\bar{\boldsymbol{S}} = \bar{\boldsymbol{A}}\boldsymbol{S}$

Let $\bar{\boldsymbol{S}} = \bar{\boldsymbol{A}}\boldsymbol{S}$, where $\boldsymbol{A}$ is a diagonal matrix, $\boldsymbol{A} = \text{diag}(z_i)$ where $z_i \in \{0, 1\}$.

Then $\bar{\boldsymbol{G}} = (\boldsymbol{S}'\boldsymbol{A}'\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{A}'\boldsymbol{W}^{-1}$

1.2.5 Other potential strategies

- Grouped version of forward/backward stepwise.
- Group regularizers: Grouped Lasso, MCP (minimax concave penalty), SCAD (smoothly clipped absolute deviation).
- Unconstrained reconciliation using in-sample observations and fitted values.