Optimal forecast reconciliation with time series selection

Xiaoqian Wang

Joint work with: Rob Hyndman, Shanika Wickramasuriya

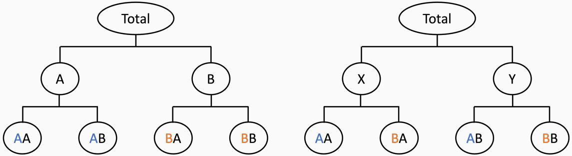
3 July 2024

Hierarchical time series

Most time series collections with linear constraints can be written as

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

- $\mathbf{y}_t$: vector of all time series at time t.
- $\mathbf{S}$: "summing matrix" containing the linear constraints.
- $\mathbf{b}_t$: vector of most disaggregated series at time t.

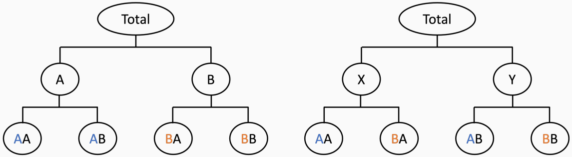


Hierarchical time series

Most time series collections with linear constraints can be written as

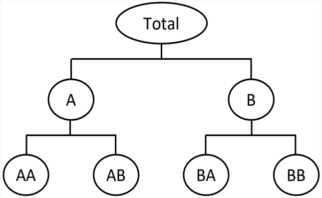
$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$$

- $\mathbf{y}_t$: vector of all time series at time t.
- $\mathbf{S}$: "summing matrix" containing the linear constraints.
- $\mathbf{b}_t$: vector of most disaggregated series at time t.

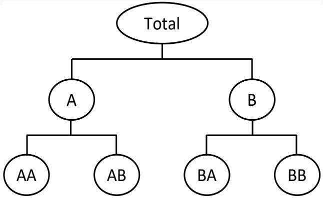


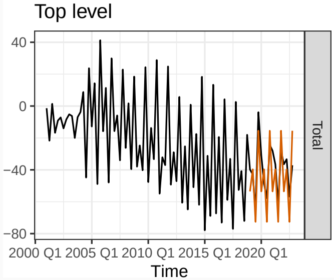
$$\mathbf{y}_t = \begin{bmatrix} y_{\text{Total},t} \\ y_{\text{A},t} \\ y_{\text{B},t} \\ y_{\text{X},t} \\ y_{\text{Y},t} \\ y_{\text{AA},t} \\ y_{\text{AB},t} \\ y_{\text{BA},t} \\ y_{\text{BB},t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ & \mathbf{I}_4 & & \end{bmatrix} \begin{bmatrix} y_{\text{AA},t} \\ y_{\text{AB},t} \\ y_{\text{BA},t} \\ y_{\text{BB},t} \end{bmatrix}$$

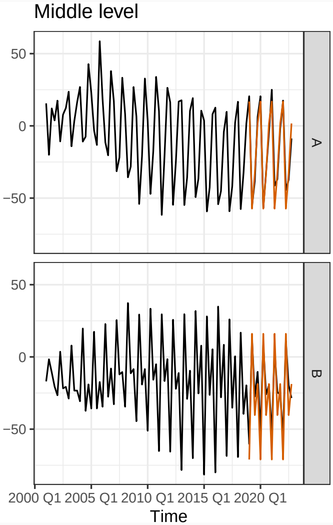
An example hierarchy

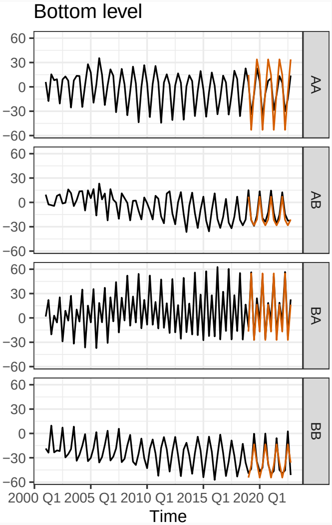


An example hierarchy









Linear forecast reconciliation

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h$$

- $\tilde{\mathbf{y}}_h$: vector of "coherent forecasts".
- $\mathbf{G}$: matrix mapping the base forecasts into bottom-level forecasts.
- $\hat{\mathbf{y}}_h$: vector of initial h-step-ahead "base forecasts" made at time t.

Linear forecast reconciliation

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h$$

- $\tilde{\mathbf{y}}_h$: vector of "coherent forecasts".
- $\mathbf{G}$: matrix mapping the base forecasts into bottom-level forecasts.
- $\hat{\mathbf{y}}_h$: vector of initial h-step-ahead "base forecasts" made at time t.

$$\begin{aligned} & \mathrm{E}\left[\|\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h\|_2^2 \mid \mathbf{I}_T\right] \\ = \quad & \underbrace{\|\mathbf{S}\mathbf{G}\left(\mathrm{E}\left[\hat{\mathbf{y}}_h \mid \mathbf{I}_T\right] - \mathrm{E}\left[\mathbf{y}_{T+h} \mid \mathbf{I}_T\right]\right) + (\mathbf{S} - \mathbf{S}\mathbf{G}\mathbf{S})\mathrm{E}\left[\mathbf{b}_{T+h} \mid \mathbf{I}_T\right]\|_2^2}_{\text{bias}} \\ & + \underbrace{\mathrm{Tr}\left(\mathrm{Var}\left[\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h \mid \mathbf{I}_T\right]\right)}_{\text{variance}} \end{aligned}$$

Linear forecast reconciliation

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h$$

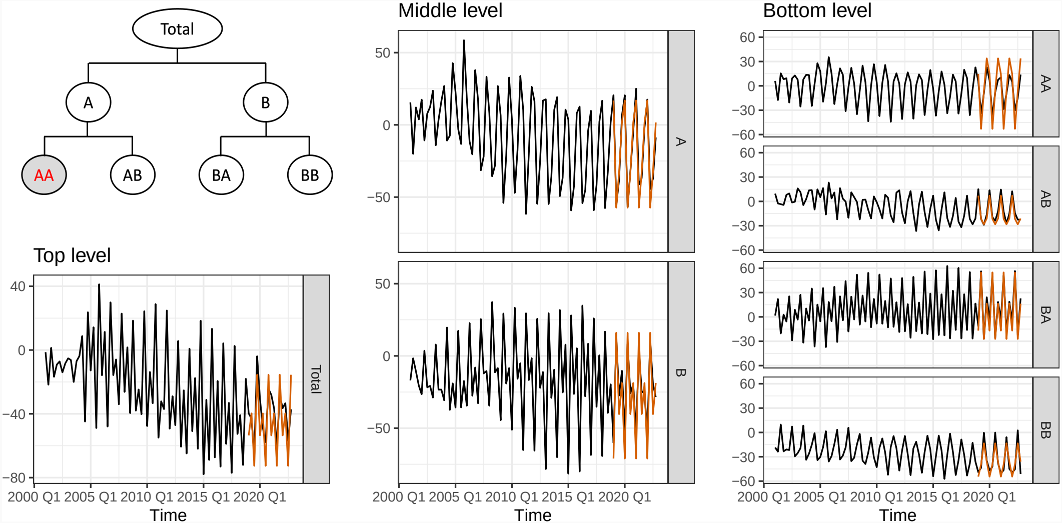
- $\tilde{\mathbf{y}}_h$: vector of ”coherent forecasts”.
- $\mathbf{G}$: matrix mapping the base forecasts into bottom-level forecasts.
- $\hat{\mathbf{y}}_h$: vector of initial h-step-ahead ”base forecasts” made at time t.

$$\begin{aligned} & \mathrm{E}\left[\|\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h\|_2^2 \mid \mathbf{I}_T\right] \\ = \quad & \underbrace{\left\|\mathbf{SG}\left(\mathrm{E}\left[\hat{\mathbf{y}}_h \mid \mathbf{I}_T\right] - \mathrm{E}\left[\mathbf{y}_{T+h} \mid \mathbf{I}_T\right]\right) + (\mathbf{S} - \mathbf{SGS})\mathrm{E}\left[\mathbf{b}_{T+h} \mid \mathbf{I}_T\right]\right\|_2^2}_{\text{bias}} \\ & + \underbrace{\mathrm{Tr}\left(\mathrm{Var}\left[\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h \mid \mathbf{I}_T\right]\right)}_{\text{variance}} \end{aligned}$$

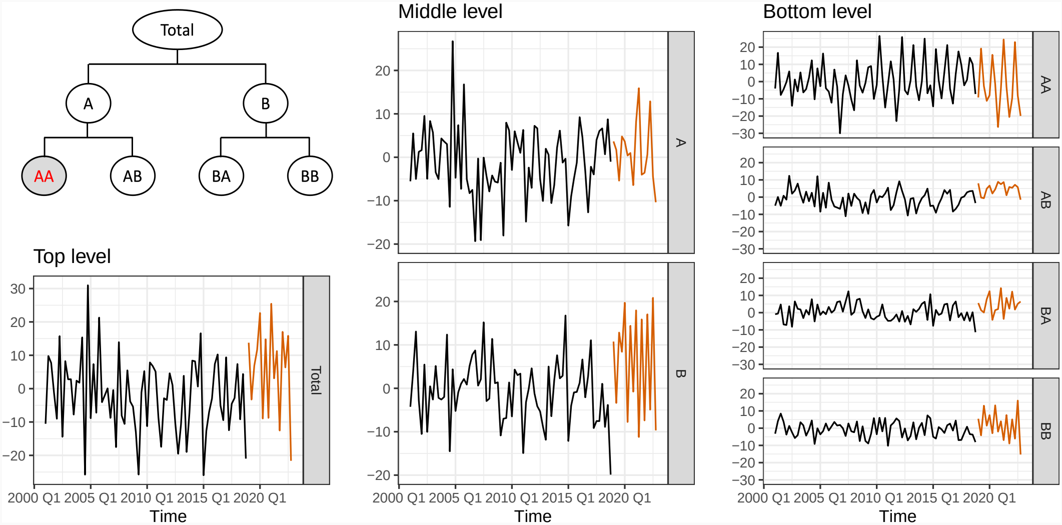
Minimum trace reconciliation (MinT)

$$\mathbf{G} = \left(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$$

The example hierarchy (observations & forecasts)



The example hierarchy (residuals & forecast errors)



How to achieve selection?

The purpose

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h$$

Eliminate the negative effect of some series on forecast reconciliation.

About $\mathbf{G}$: Zero out some columns of $\mathbf{G}$.

About $\mathbf{S}$: Do not zero out the corresponding rows of $\mathbf{S}$.

How to achieve selection?

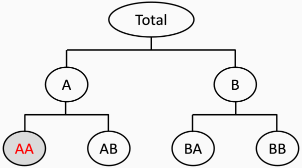
The purpose

$$\tilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h$$

Eliminate the negative effect of some series on forecast reconciliation.

About $\mathbf{G}$: Zero out some columns of $\mathbf{G}$.

About $\mathbf{S}$: Do not zero out the corresponding rows of $\mathbf{S}$.



$$\begin{bmatrix} \tilde{y}_{\text{Total}} \\ \tilde{y}_{\text{A}} \\ \tilde{y}_{\text{B}} \\ \tilde{y}_{\text{AA}} \\ \tilde{y}_{\text{AB}} \\ \tilde{y}_{\text{BA}} \\ \tilde{y}_{\text{BB}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & w_{13} & 0 & w_{15} & w_{16} & w_{17} \\ w_{21} & w_{22} & w_{23} & 0 & w_{25} & w_{26} & w_{27} \\ w_{31} & w_{32} & w_{33} & 0 & w_{35} & w_{36} & w_{37} \\ w_{41} & w_{42} & w_{43} & 0 & w_{45} & w_{46} & w_{47} \end{bmatrix} \begin{bmatrix} \hat{y}_{\text{Total}} \\ \hat{y}_{\text{A}} \\ \hat{y}_{\text{B}} \\ \hat{y}_{\text{AA}} \\ \hat{y}_{\text{AB}} \\ \hat{y}_{\text{BA}} \\ \hat{y}_{\text{BB}} \end{bmatrix}$$

Method I: Regularized best-subset selection

Group best-subset selection with ridge regularization

$$\min_{\mathbf{G}} \quad \frac{1}{2}\left(\hat{\mathbf{y}} - \mathbf{SG}\hat{\mathbf{y}}\right)' \mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{SG}\hat{\mathbf{y}}\right) + \lambda_0 \sum_{j=1}^{n} \mathbf{1}\left(\mathbf{G}_{\cdot j} \neq \mathbf{0}\right) + \lambda_2 \left\|\operatorname{vec}\left(\mathbf{G}\right)\right\|_2^2$$

$$\text{s.t.} \quad \mathbf{GS} = \mathbf{I}_{n_b}$$

- $\mathbf{1}(\cdot)$: the indicator function.
- $\lambda_0 > 0$: controls the number of nonzero columns of G selected.
- $\lambda_2 \geq 0$: controls the strength of the ridge regularization.

Method I: Regularized best-subset selection

- $\mathbf{SG\hat{y}} = \mathrm{vec}\left(\mathbf{SG\hat{y}}\right) = \left(\mathbf{\hat{y}}' \otimes \mathbf{S}\right)\mathrm{vec}(\mathbf{G})$.

Big-M based MIP formulation (MIQP)

$$\min_{\mathbf{G},\mathbf{z},\check{\mathbf{e}},\mathbf{g}^+} \frac{1}{2}\check{\mathbf{e}}'\mathbf{W}^{-1}\check{\mathbf{e}} + \lambda_0 \sum_{j=1}^{n} z_j + \lambda_2 \mathbf{g}^{+\prime}\mathbf{g}^+$$

s.t.

$$\mathbf{GS} = \mathbf{I}_{n_b} \Leftrightarrow \left(\mathbf{S}' \otimes \mathbf{I}_{n_b}\right)\mathrm{vec}(\mathbf{G}) = \mathrm{vec}\left(\mathbf{I}_{n_b}\right) \quad \cdots (C1)$$

$$\mathbf{\hat{y}} - \left(\mathbf{\hat{y}}' \otimes \mathbf{S}\right)\mathrm{vec}(\mathbf{G}) = \check{\mathbf{e}} \quad \cdots (C2)$$

$$\sum_{i=1}^{n_b} g^+_{i+(j-1)n_b} \leqslant \mathcal{M} z_j, \quad j \in [n] \quad \cdots (C3)$$

$$\mathbf{g}^+ \geqslant \mathrm{vec}(\mathbf{G}) \quad \cdots (C4)$$

$$\mathbf{g}^+ \geqslant -\mathrm{vec}(\mathbf{G}) \quad \cdots (C5)$$

$$z_j \in \{0, 1\}, \quad j \in [n] \quad \cdots (C6)$$

Method II: Intuitive method

The MinT solution: $\mathbf{G} = \left(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$.

Based on MinT solution, we assume $\bar{\mathbf{G}} = (\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1}\mathbf{AS})^{-1}\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1}$.

- $\bar{\mathbf{S}} = \mathbf{AS}$.
- $\boldsymbol{A} = \mathrm{diag}(\boldsymbol{z})$ is a diagonal matrix with $z_j \in \{0, 1\}$ for $j \in [n]$.
- Estimate the whole $\mathbf{G} \Longrightarrow$ estimate $\mathbf{A}$.

Method II: Intuitive method

The MinT solution: $\mathbf{G} = \left(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{W}_h^{-1}$.

Based on MinT solution, we assume $\bar{\mathbf{G}} = (\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1}\mathbf{A}\mathbf{S})^{-1}\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1}$.

- $\bar{\mathbf{S}} = \mathbf{A}\mathbf{S}$.
- $\mathbf{A} = \mathrm{diag}(\mathbf{z})$ is a diagonal matrix with $z_j \in \{0, 1\}$ for $j \in [n]$.
- Estimate the whole $\mathbf{G} \Longrightarrow$ estimate $\mathbf{A}$.

Intuitive method with L_0 regularization

$$\min_{\mathbf{A}} \quad \frac{1}{2}\left(\hat{\mathbf{y}} - \mathbf{S}\bar{\mathbf{G}}\hat{\mathbf{y}}\right)'\mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{S}\bar{\mathbf{G}}\hat{\mathbf{y}}\right) + \lambda_0 \sum_{j=1}^{n} \mathbf{A}_{jj}$$

$$\text{s.t.} \quad \bar{\mathbf{G}} = (\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1}\mathbf{A}\mathbf{S})^{-1}\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1}$$

$$\bar{\mathbf{G}}\mathbf{S} = \mathbf{I}$$

Method II: Intuitive method

Toy example

```
S <- rbind(c(1,1,1,1), c(1,1,0,0), c(0,0,1,1), diag(1,4))
W_inv <- diag(c(4,2,2,rep(1,4))) |> solve()
G <- solve(t(S) %*% W_inv %*% S) %*% (t(S) %*% W_inv) |> round(2)

A <- diag(c(1,0,rep(1, 5)))
G_bar <- solve(t(A %*% S) %*% W_inv %*% A %*% S) %*% (t(A %*% S) %*% W_inv) |> round(2)
list(G = G, G_bar = G_bar)
```

```
$G
     [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
[1,] 0.08  0.21 -0.04  0.71 -0.29 -0.04 -0.04
[2,] 0.08  0.21 -0.04 -0.29  0.71 -0.04 -0.04
[3,] 0.08 -0.04  0.21 -0.04 -0.04  0.71 -0.29
[4,] 0.08 -0.04  0.21 -0.04 -0.04 -0.29  0.71

$G_bar
     [,1] [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
[1,] 0.14    0 -0.07  0.86 -0.14 -0.07 -0.07
[2,] 0.14    0 -0.07 -0.14  0.86 -0.07 -0.07
[3,] 0.07    0  0.21 -0.07 -0.07  0.71 -0.29
[4,] 0.07    0  0.21 -0.07 -0.07 -0.29  0.71
```

Method II: Intuitive method

Toy example

```
S <- rbind(c(1,1,1,1), c(1,1,0,0), c(0,0,1,1), diag(1,4))
W_inv <- diag(c(4,2,2,rep(1,4))) |> solve()
G <- solve(t(S) %*% W_inv %*% S) %*% (t(S) %*% W_inv) |> round(2)

A <- diag(c(1,0,rep(1, 5)))
G_bar <- solve(t(A %*% S) %*% W_inv %*% A %*% S) %*% (t(A %*% S) %*% W_inv) |> round(2)
list(G = G, G_bar = G_bar)
```

MIP formulation (MIQP)

$$
\begin{aligned}
\min_{\mathbf{A},\bar{\mathbf{G}},\mathbf{C},\check{\mathbf{e}},\mathbf{z}} \quad & \frac{1}{2}\check{\mathbf{e}}'\mathbf{W}^{-1}\check{\mathbf{e}} + \lambda_0 \sum_{j=1}^{n} z_j \\
\text{s.t.} \quad & \bar{\mathbf{G}}\mathbf{S} = \mathbf{I} \\
& \hat{\mathbf{y}} - \left(\hat{\mathbf{y}}' \otimes \mathbf{S}\right) \operatorname{vec}(\bar{\mathbf{G}}) = \check{\mathbf{e}} \\
& \bar{\mathbf{G}}\mathbf{A}\mathbf{S} = \mathbf{I} \\
& \bar{\mathbf{G}} = \mathbf{C}\mathbf{S}'\mathbf{A}'\mathbf{W}^{-1} \\
& z_j \in \{0,1\}, \quad j \in [n]
\end{aligned}
$$

Method III: Group lasso method

Group lasso with the unbiasedness constraint

$$\min_{\mathbf{G}} \quad \frac{1}{2}\left(\hat{\mathbf{y}} - \mathbf{SG}\hat{\mathbf{y}}\right)' \mathbf{W}^{-1}\left(\hat{\mathbf{y}} - \mathbf{SG}\hat{\mathbf{y}}\right) + \lambda \sum_{j=1}^{n} w_j \left\|\mathbf{G}_{\cdot j}\right\|_2$$

$$\text{s.t.} \quad \mathbf{GS} = \mathbf{I}_{n_b}$$

- $\lambda \geq 0$: tuning parameter.
- $w_j \neq 0$: penalty weight in order to make model more flexible.

Method III: Group lasso method

Second order cone programming formulation (SOCP)

$$\min_{\mathbf{G},\check{\mathbf{e}},\mathbf{g}^+} \frac{1}{2}\check{\mathbf{e}}'\mathbf{W}_h^{-1}\check{\mathbf{e}} + \lambda \sum_{j=1}^{n} w_j c_j$$

s.t.

$$\left(\mathbf{S}' \otimes \mathbf{I}_{n_b}\right) \mathrm{vec}(\mathbf{G}) = \mathrm{vec}\left(\mathbf{I}_{n_b}\right)$$

$$\hat{\mathbf{y}} - (\hat{\mathbf{y}}' \otimes \mathbf{S})\,\mathrm{vec}(\mathbf{G}) = \check{\mathbf{e}}$$

$$c_j = \sqrt{\sum_{i=1}^{n_b} g^{+2}_{i+(j-1)n_b}}, \quad j \in [n]$$

Proposition 1

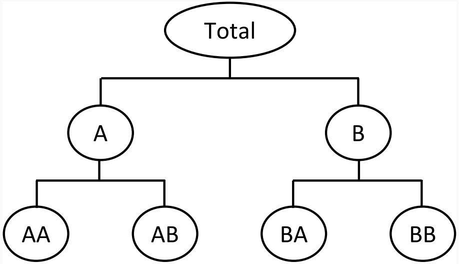
Proposition 1

- If the assumption that forecast reconciliation preserves unbiasedness is imposed by enforcing $\mathbf{GS} = \mathbf{I}$, then the number of nonzero column entries of $\hat{\mathbf{G}}$ will be no less than n_b.
- The constraint $\mathbf{GS} = \mathbf{I}$ enforces that the selected columns of $\hat{\mathbf{G}}$ will correspond to variables that can "restore" the hierarchy.

Proposition 1

Proposition 1

- If the assumption that forecast reconciliation preserves unbiasedness is imposed by enforcing $\mathbf{GS} = \mathbf{I}$, then the number of nonzero column entries of $\hat{\mathbf{G}}$ will be no less than n_b.
- The constraint $\mathbf{GS} = \mathbf{I}$ enforces that the selected columns of $\hat{\mathbf{G}}$ will correspond to variables that can ”restore” the hierarchy.



Method IV: Empirical group lasso method

Empirical group lasso

$$\min_{\mathbf{G}} \quad \frac{1}{2T}\left\|\mathbf{Y} - \hat{\mathbf{Y}}\mathbf{G}'\mathbf{S}'\right\|_F^2 + \lambda \sum_{j=1}^{n} w_j \left\|\mathbf{G}_{\cdot j}\right\|_2$$

- $\mathbf{Y} \in \mathbb{R}^{T\times n}$: a matrix comprising observations on the training set.
- $\hat{\mathbf{Y}} \in \mathbb{R}^{T\times n}$: a matrix of in-sample one-step-ahead forecasts.
- $\lambda \geq 0$: a tuning parameter.
- $w_j \neq 0$: penalty weight assigned in $\mathbf{G}_{\cdot j}$.

Method IV: Empirical group lasso method

Empirical group lasso

$$\min_{\mathbf{G}} \quad \frac{1}{2T}\left\|\mathbf{Y}-\hat{\mathbf{Y}}\mathbf{G}'\mathbf{S}'\right\|_F^2 + \lambda\sum_{j=1}^{n} w_j \left\|\mathbf{G}_{\cdot j}\right\|_2$$

- $\mathbf{Y} \in \mathbb{R}^{T\times n}$: a matrix comprising observations on the training set.
- $\hat{\mathbf{Y}} \in \mathbb{R}^{T\times n}$: a matrix of in-sample one-step-ahead forecasts.
- $\lambda \geq 0$: a tuning parameter.
- $w_j \neq 0$: penalty weight assigned in $\mathbf{G}_{\cdot j}$.

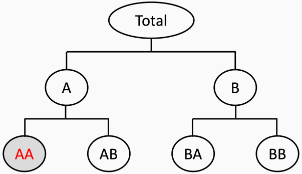
Standard group lasso problem

$$\min_{\mathrm{vec}(\mathbf{G})} \quad \frac{1}{2T}\left\|\mathrm{vec}(\mathbf{Y}) - (\mathbf{S}\otimes\hat{\mathbf{Y}})\,\mathrm{vec}\left(\mathbf{G}'\right)\right\|_2^2 + \lambda\sum_{j=1}^{n} w_j \left\|\mathbf{G}_{\cdot j}\right\|_2$$

Simulation setup

Data generation

Bottom-level series:



$$\mathbf{b}_t = \mu_t + \gamma_t + \eta_t$$

where

$$\mu_t = \mu_{t-1} + \upsilon_t + \varrho_t, \qquad \varrho_t \sim N\left(0, \sigma_\varrho^2 I_4\right),$$

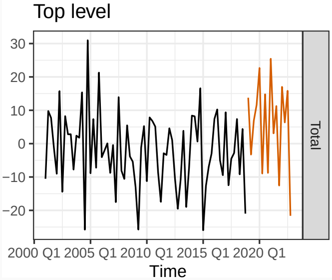
$$\upsilon_t = \upsilon_{t-1} + \zeta_t, \qquad \zeta_t \sim \mathcal{N}\left(0, \sigma_\zeta^2 I_4\right),$$

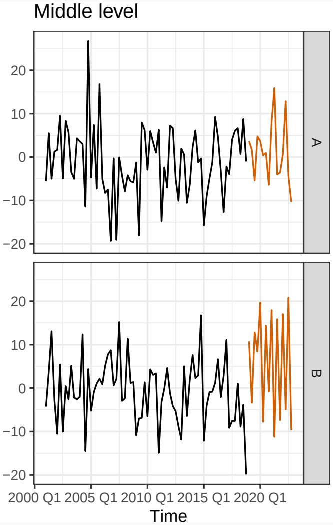
$$\gamma_t = -\sum_{i=1}^{s-1} \gamma_{t-i} + \omega_t, \quad \omega_t \sim \mathcal{N}\left(0, \sigma_\omega^2 I_4\right),$$

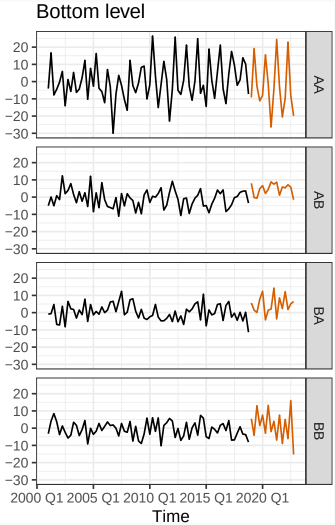
and ϱ_t, ζ_t, and ω_t are errors independent of each other and over time.

Results for the simple example (residuals & forecast errors)

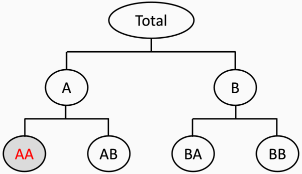
	Top	A	B	AA	AB	BA	BB
OLS_subset	1	1	1	1	1	1	1
WLSs_subset	1	1	1	1	1	1	1
WLSv_subset	1	0	0	0	1	1	1
MinT_subset	1	0	1	0	1	1	0
MinTs_subset	1	0	1	0	1	1	0







Results



- A - AB
- Total - B - AB
- Total - BA - BB - AB

Proportion of time series being selected (AA is deteriorated).

	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	

Results

Out-of-sample forecast results (RMSE) for the simulated data (AA is 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**
EMinT	2.2	2.9	2.5	1.7	2.5	2.9	2.3	1.3	–31.9	–32.3	–27.5	–22.0	–15.9	–16.2	–13.4	–10.5
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

Summary

Key takeaways

- Exclude poorly performing base forecasts when performing reconciliation.
- Reduce disparities from using different estimates of $\mathbf{W}$.
- Demonstrate effectiveness in addressing model misspecification issues.
- Perform better or comparably than benchmarks when no model misspecification is apparent.

Limitations

- Addressing L_0-regularized regression problems with additional constraints remains challenging.
- Introducing a bias correction when the unbiasedness preserving property is dropped.

More information

- **Paper and code:** xqnwang.rbind.io/publication/hfs
- **Slides:** xqnwang.rbind.io/talk/isf2024

Find me at ...

xqnwang.rbind.io

Xia0qianWang

@xqnwang

xiaoqian.wang@monash.edu