MONASH University

OPTiMA ARC TRAINING CENTRE IN OPTIMISATION TECHNOLOGIES INTEGRATED METHODOLOGIES AND APPLICATIONS

MONASH BUSINESS SCHOOL

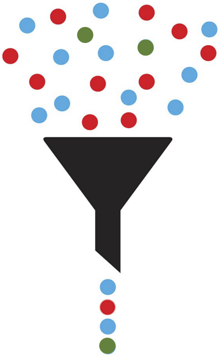
Optimal forecast reconciliation with time series selection

Xiaoqian Wang

In collaboration with: Rob Hyndman & Shanika Wickramasuriya

Department of Econometrics and Business Statistics

MONASH University



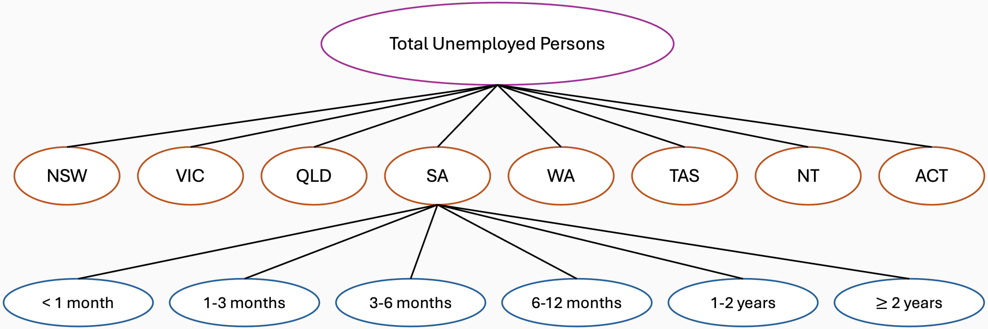
Outline

Outline

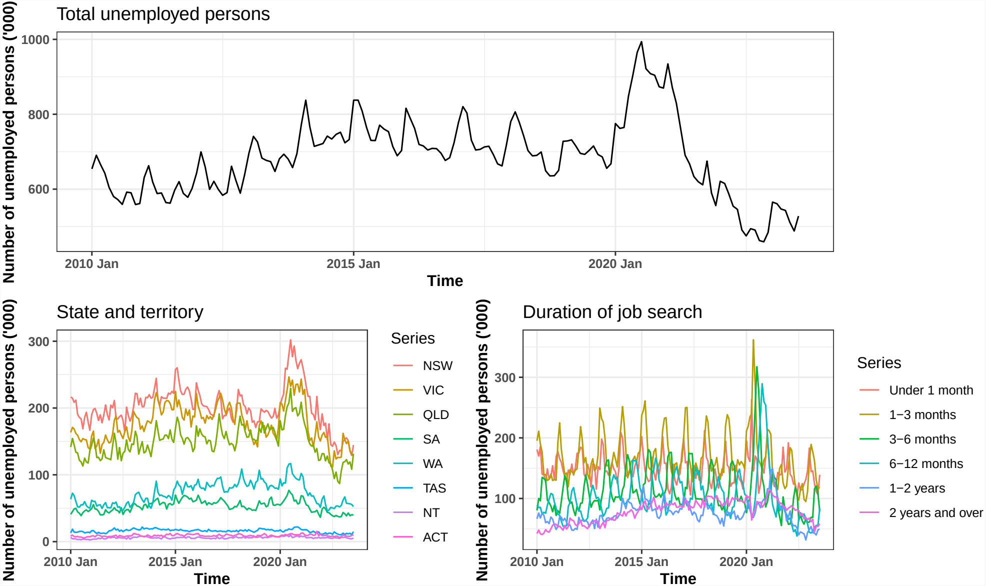
Australian labour force data

Total number of unemployed persons in Australia

- Eight states and territories
  - Six different groups of job search duration

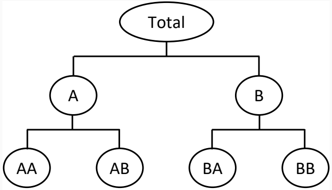


Australian labour force data

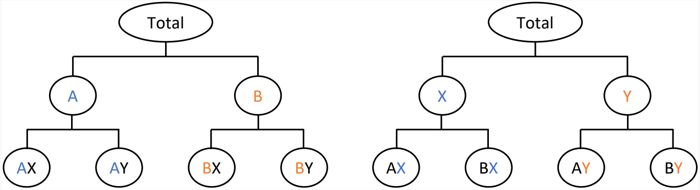


Hierarchical and grouped time series

A **hierarchical time series** is a collection of several time series that are linked together in a hierarchical structure.



A **grouped time series** does not naturally aggregate (or disaggregate) in a unique manner.

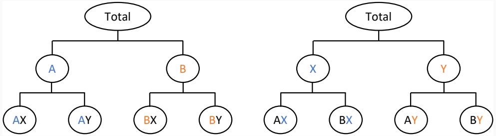


Notation

Almost all collections of time series with linear constraints can be written as

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

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



$$\boldsymbol{y}_t = \begin{bmatrix} y_{\text{Total},t} \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ & \boldsymbol{I}_4 & & \end{bmatrix} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{bmatrix}$$

Coherence

Coherence is the property that data adhere to the linear constraints.

Hierarchical forecasting problem

- Observations naturally adhere to these linear constraints.
- We can use any (independent) method to generate forecasts of all series, but in general they will not be coherent.
- Constraints should be imposed on the forecasts to ensure coherence.

Forecast reconciliation is a post-processing method that ensures forecasts of multivariate time series adhere to known linear constraints.

Outline

1. Hierarchical time series
2. Linear forecast reconciliation
3. Forecast reconciliation with time series selection
4. Simulation experiments
5. Forecasting Australian labour force
6. Conclusions

Linear forecast reconciliation

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

- $\hat{\boldsymbol{y}}_h$: vector of initial h-step-ahead "base forecasts" made at time T.
- $\boldsymbol{G}$: matrix combining all base forecasts to form bottom-level reconciled forecasts.
- $\boldsymbol{S}$: summing matrix containing the linear constraints.
- $\tilde{\boldsymbol{y}}_h$: vector of "coherent forecasts".

Single-level approaches

- Bottom-Up: $\boldsymbol{G}_{BU} = \left[\boldsymbol{O}_{n_b \times n_a} \mid \boldsymbol{I}_{n_b}\right]$.
- Top-Down: $\boldsymbol{G}_{TD} = \left[\boldsymbol{p} \mid \boldsymbol{O}_{n_b \times (n-1)}\right]$ and $\sum_{i=1}^{n_b} p_i = 1$.

Minimum trace reconciliation

- Problem: minimizing the trace of the covariance matrix $\mathrm{Var}\left(\boldsymbol{y}_h - \tilde{\boldsymbol{y}}_h\right)$.
- Solution: $\boldsymbol{G} = \left(\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}$.
- $\boldsymbol{W}_h$ estimators: OLS, WLSs, WLSv, MinT, MinTs.

Linear forecast reconciliation

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

- $\hat{\boldsymbol{y}}_h$: vector of initial h-step-ahead "base forecasts" made at time T.
- $\boldsymbol{G}$: matrix combining all base forecasts to form bottom-level reconciled forecasts.
- $\boldsymbol{S}$: summing matrix containing the linear constraints.
- $\tilde{\boldsymbol{y}}_h$: vector of "coherent forecasts".

Different estimators of $\boldsymbol{W}$

Reconciliation method	$\boldsymbol{W}_h \propto$
OLS (Hyndman et al. 2011)	$\boldsymbol{I}$
WLSs (Athanasopoulos et al. 2017)	$\mathrm{diag}(\boldsymbol{S}1)$
WLSv (Hyndman et al. 2016)	$\mathrm{Diag}\left(\hat{\boldsymbol{W}}_1\right)$
MinT (Wickramasuriya et al. 2019)	$\hat{\boldsymbol{W}}_1$
MinTs (Wickramasuriya et al. 2019)	$\lambda\,\mathrm{Diag}\left(\hat{\boldsymbol{W}}_1\right) + (1-\lambda)\hat{\boldsymbol{W}}_1$

Intuition behind $\boldsymbol{W}$

The trace minimization problem can be reformulated as a linear equality constrained least squares problem.

Optimization problem

$$\min_{\tilde{\boldsymbol{y}}} \quad \frac{1}{2}(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})'\boldsymbol{W}^{-1}(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}})$$
$$\text{s.t.} \quad \tilde{\boldsymbol{y}} = \boldsymbol{S}\tilde{\boldsymbol{b}}$$

- Generalized Least Squares problem.
- The larger the estimated variance of the base forecast errors, the larger the range of adjustments permitted for forecast reconciliation.

Some potential issues

- Disparities emerge due to the use of different estimates of $\boldsymbol{W}$, making it challenging to choose the "right" estimator.
- Some series may experience deteriorations in reconciled forecasts, especially those with poor forecasts.
- The lack of use of in-sample information.

Outline

How to achieve selection?

The purpose

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

Eliminate the negative effect of some series on forecast reconciliation.

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

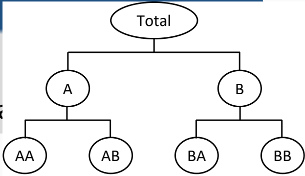
About $\boldsymbol{S}$: Do not zero out the corresponding rows of $\boldsymbol{S}$.

How to achieve selection?

The purpose

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{SG}\hat{\boldsymbol{y}}_h$$

Eliminate the negative effect of some series on forecast reconcilia



About *G*: Zero out some columns of ***G***.

About *S*: Do not zero out the corresponding rows of ***S***.

$$\begin{bmatrix} \tilde{y}_{\text{Total}} \\ \tilde{y}_{\text{A}} \\ \tilde{y}_{\text{B}} \\ \tilde{y}_{\text{AA}} \\ \tilde{y}_{\text{AB}} \\ \tilde{y}_{\text{BA}} \\ \tilde{y}_{\text{BB}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_{11} & 0 & w_{13} & w_{14} & w_{15} & w_{16} & w_{17} \\ w_{21} & 0 & w_{23} & w_{24} & w_{25} & w_{26} & w_{27} \\ w_{31} & 0 & w_{33} & w_{34} & w_{35} & w_{36} & w_{37} \\ w_{41} & 0 & w_{43} & w_{44} & w_{45} & w_{46} & w_{47} \end{bmatrix} \begin{bmatrix} \hat{y}_{\text{Total}} \\ \hat{y}_{\text{A}} \\ \hat{y}_{\text{B}} \\ \hat{y}_{\text{AA}} \\ \hat{y}_{\text{AB}} \\ \hat{y}_{\text{BA}} \\ \hat{y}_{\text{BB}} \end{bmatrix}$$

Method I: Regularized best-subset selection

Group best-subset selection with ridge regularization

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

$$\text{s.t.} \quad \boldsymbol{G}\boldsymbol{S} = \boldsymbol{I}_{n_b}$$

- $\mathbf{1}(\cdot)$: the indicator function.
- $\lambda_0 > 0$: controls the number of nonzero columns of $\boldsymbol{G}$ selected.
- $\lambda_2 \geq 0$: controls the strength of the ridge regularization.
- $\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}} = \mathrm{vec}\left(\boldsymbol{S}\boldsymbol{G}\hat{\boldsymbol{y}}\right) = \left(\hat{\boldsymbol{y}}' \otimes \boldsymbol{S}\right) \mathrm{vec}(\boldsymbol{G})$.

Method I: Regularized best-subset selection

Big-M based MIP formulation (MIQP)

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

$$\text{s.t.} \quad \boldsymbol{G}\boldsymbol{S} = \boldsymbol{I}_{n_b} \Leftrightarrow \left(\boldsymbol{S}' \otimes \boldsymbol{I}_{n_b}\right) \mathrm{vec}(\boldsymbol{G}) = \mathrm{vec}\left(\boldsymbol{I}_{n_b}\right) \quad \cdots (C1)$$

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

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

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

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

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

Method I: Regularized best-subset selection

Hyperparameter

- ℓ_0 regularization parameter
  - $\lambda_0^1 = \frac{1}{2}\left(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}^{\text{bench}}\right)' \boldsymbol{W}^{-1}\left(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}^{\text{bench}}\right)$
  - $\lambda_0^k = 0.0001\lambda_0^1$
  - Generate a grid of $k+1$ values, $\lambda_0 = \{\lambda_0^1, \ldots, \lambda_0^k, 0\}$, where $\lambda_0^j = \lambda_0^1\left(\lambda_0^k/\lambda_0^1\right)^{(j-1)/(k-1)}$ for $j \in [k]$.
- ℓ_2 regularization parameter
  - $\lambda_2 = \{0, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$
- Tune the parameters to minimize the sum of squared reconciled forecast errors on a truncated training set.

Method II: Intuitive method

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

We utilize the MinT solution and assume $\bar{\boldsymbol{G}} = (\boldsymbol{S}'\boldsymbol{A}'\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{A}'\boldsymbol{W}^{-1}$.

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

Intuitive method with L_0 regularization

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

Method II: Intuitive method

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

MIP formulation (MIQP)

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

Method II: Intuitive method

MIP formulation (MIQP)

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

Hyperparameter (ℓ_0 regularization parameter)

- $\lambda_0^1 = \frac{1}{2}\left(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}^{\text{bench}}\right)' \boldsymbol{W}^{-1} \left(\hat{\boldsymbol{y}} - \tilde{\boldsymbol{y}}^{\text{bench}}\right)$, $\lambda_0^k = 0.0001\lambda_0^1$.
- $\lambda_0 = \{\lambda_0^1, \ldots, \lambda_0^k, 0\}$, where $\lambda_0^j = \lambda_0^1 \left(\lambda_0^k / \lambda_0^1\right)^{(j-1)/(k-1)}$ for $j \in [k]$.

Method III: Group lasso method

Group lasso with the unbiasedness constraint

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

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

- $\lambda \geq 0$: 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)

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

$$\text{s.t.} \quad \left(\boldsymbol{S}' \otimes \boldsymbol{I}_{n_b}\right) \operatorname{vec}(\boldsymbol{G}) = \operatorname{vec}\left(\boldsymbol{I}_{n_b}\right)$$

$$\hat{\boldsymbol{y}} - \left(\hat{\boldsymbol{y}}' \otimes \boldsymbol{S}\right) \operatorname{vec}(\boldsymbol{G}) = \check{\boldsymbol{e}}$$

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

Method III: Group lasso method

Hyperparameter

- Penalty weights: $w_j = 1/\left\|\boldsymbol{G}_{\cdot j}^{\text{bench}}\right\|_2$.
- λ sequence.
  - Ignoring the unbiasedness constraint,

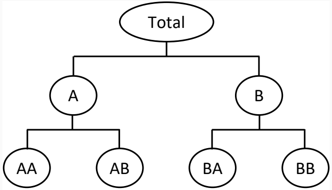
$$\lambda^1 = \max_{j=1,\ldots,n} \left\| -\left((\hat{\boldsymbol{y}}' \otimes \boldsymbol{S})_{\cdot j*}\right)' \boldsymbol{W}^{-1}\hat{\boldsymbol{y}}\right\|_2 / w_j$$

    is the smallest λ value such that all predictors have zero coefficients, i.e., $\boldsymbol{G} = \boldsymbol{O}$.
  - $\lambda^k = 0.0001\lambda^1$.
  - $\lambda = \{\lambda^1, \ldots, \lambda^k, 0\}$, where $\lambda^j = \lambda^1(\lambda^k/\lambda^1)^{(j-1)/(k-1)}$ for $j \in [k]$.

Proposition 1

Proposition 1

Under the assumption of unbiasedness, the count of nonzero column entries of $\boldsymbol{G}$ (i.e., the number of time series selected for reconciliation) is at least equal to the number of time series at the bottom level. In addition, we can restore the full hierarchical structure by aggregating/disaggregating the selected time series.



Method IV: Empirical group lasso method

Empirical group lasso

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

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

Method IV: Empirical group lasso method

Empirical group lasso

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

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

Standard group lasso problem

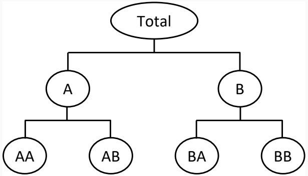
$$\min_{\mathrm{vec}(\boldsymbol{G})} \quad \frac{1}{2T}\left\|\mathrm{vec}(\boldsymbol{Y}) - (\boldsymbol{S} \otimes \hat{\boldsymbol{Y}})\,\mathrm{vec}\left(\boldsymbol{G}'\right)\right\|_2^2 + \lambda \sum_{j=1}^{n} w_j \left\|\boldsymbol{G}_{\cdot j}\right\|_2$$

Outline

Setup 1: Exploring the effect of model misspecification

Data generation

Bottom-level series:



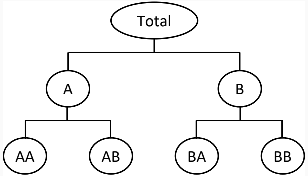
$$\boldsymbol{b}_t = \boldsymbol{\mu}_t + \boldsymbol{\gamma}_t + \boldsymbol{\eta}_t$$

where

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

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

Setup 1: Exploring the effect of model misspecification



Other details

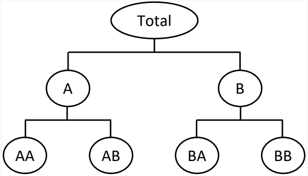
- $s = 4$ for quarterly data, $T + h = 180$, $h = 16$.
- $\sigma_{\varrho}^2 = 2, \sigma_{\zeta}^2 = 0.007$, and $\sigma_{\omega}^2 = 7$.
- Initial values for $\boldsymbol{\mu}_0, \boldsymbol{v}_0, \gamma_0, \gamma_1, \gamma_2$ were generated independently from a multivariate normal distribution with mean zero and covariance matrix, $\Sigma_0 = I_4$.
- $\boldsymbol{\eta}_t$ is generated independently from an ARIMA$(p, 0, q)$ process, where p and q take values of 0 or 1 with equal probability.
- The bottom-level series are aggregated for data at higher levels.
- This process was repeated 500 times.
- Base forecasts are generated using ETS models.

Results (series A at the middle level is undermined)

Out-of-sample forecast results (RMSE) for the simulated data in Scenario B, Setup 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	9.6	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**
EMinT	2.2	2.9	2.5	1.7	–46.2	–48.1	–42.4	–35.3	3.6	2.9	2.0	1.1	–20.5	–21.9	–18.2	–14.3
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

Results (series A at the middle level is undermined)

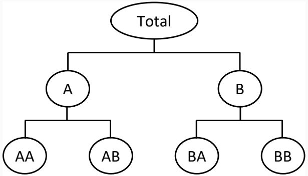


Proportion of time series being selected in Scenario B, Setup 1.

	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	

Setup 2: Exploring the effect of correlation

Data generation



Bottom-level series VAR(1):

$$\boldsymbol{b}_t = \boldsymbol{c} + \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix} \boldsymbol{b}_{t-1} + \varepsilon_t,$$

where $\boldsymbol{c}$ is a constant vector with all entries set to 1, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are 2×2 matrices with eigenvalues $z_{1,2} = 0.6[\cos(\pi/3) \pm i \sin(\pi/3)]$ and $z_{3,4} = 0.9[\cos(\pi/6) \pm i \sin(\pi/6)]$, respectively, $\varepsilon_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where

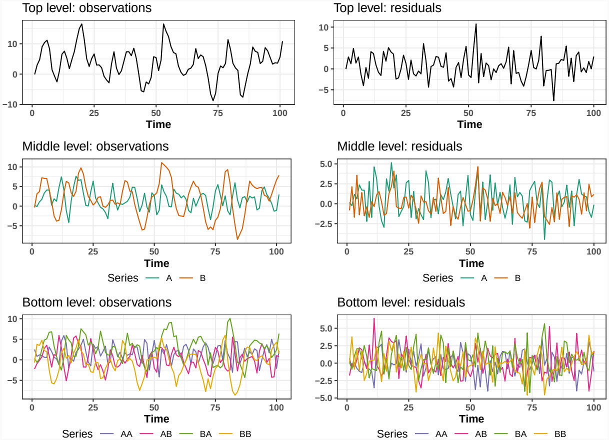
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & \boldsymbol{\Sigma}_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & \sqrt{6}\rho \\ \sqrt{6}\rho & 3 \end{bmatrix},$$

and $\rho \in \{0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8\}$.

- $T + h = 101$, $h = 1$.
- The bottom-level series are aggregated for data at higher levels.
- The process is repeated 500 times for each candidate correlation, ρ.

Setup 2: Exploring the effect of correlation

- Base forecasts are generated using ARMA models.
- The constant term is omitted (Total, A, and BA) to introduce a slight bias.



Results

Out-of-sample forecast results (RMSE) across various error correlations for simulation in Setup 2.

	Top					Middle					Bottom					Average				
Method	ρ=–0.8	–0.4	0	0.4	0.8	ρ=–0.8	–0.4	0	0.4	0.8	ρ=–0.8	–0.4	0	0.4	0.8	ρ=–0.8	–0.4	0	0.4	0.8
Base	2.4	2.9	3.4	4.1	4.0	1.5	1.8	2.1	2.4	2.5	1.5	1.5	1.5	1.5	1.4	1.6	1.8	2.0	2.1	2.1
BU	–17.0	–9.0	–6.7	–7.0	–7.4	–6.8	0.4	4.8	5.7	2.8	0.0	0.0	0.0	0.0	0.0	–5.3	–1.9	–0.2	–0.1	–1.0
OLS	–11.0	–8.2	–7.7	–8.2	–8.0	–3.5	–0.7	3.1	2.5	0.8	0.7	–0.6	–2.0	–2.3	–2.1	–2.8	–2.4	–1.8	–2.4	–2.7
OLS-subset	**–11.4**	**–8.4**	**–8.1**	**–8.4**	**–8.8**	**–3.7**	–0.7	3.2	2.5	**0.4**	**0.3**	**–0.8**	–2.0	–1.7	**–2.6**	**–3.2**	**–2.5**	**–1.9**	–2.2	**–3.2**
OLS-intuitive	**–11.6**	–8.0	**–7.8**	–8.0	**–8.4**	**–3.6**	–0.4	3.7	2.5	**0.3**	**0.6**	–0.2	–1.3	–0.4	–1.5	**–3.0**	–2.0	–1.3	–1.6	**–2.8**
OLS-lasso	**–19.2**	**–9.8**	–7.2	**–8.7**	**–8.2**	**–10.5**	**–1.7**	**2.9**	**2.4**	0.8	**–0.8**	**–0.8**	–1.6	–2.3	–2.1	**–7.1**	**–3.1**	–1.6	**–2.5**	**–2.8**
WLSs	–16.8	–11.1	–9.6	–10.4	–10.2	–8.1	–2.8	1.5	1.2	–0.4	–0.3	–1.1	–2.4	–2.9	–2.9	–5.7	–3.9	–3.0	–3.6	–4.0
WLSs-subset	**–17.3**	**–11.4**	**–9.9**	**–11.1**	**–10.8**	**–8.3**	–2.8	**1.4**	**0.7**	**–0.9**	**–0.7**	**–1.3**	–2.4	**–3.2**	**–3.3**	**–6.1**	**–4.0**	**–3.1**	**–4.1**	**–4.5**
WLSs-intuitive	**–16.9**	**–11.5**	**–9.8**	–10.0	**–10.6**	**–8.5**	–2.8	**1.4**	1.5	**–0.7**	**–0.7**	**–1.2**	–2.3	–2.7	**–3.0**	**–6.1**	**–4.0**	–3.0	–3.3	**–4.3**
WLSs-lasso	**–18.3**	–11.1	–9.2	**–10.5**	–9.8	**–9.3**	–2.4	**1.4**	1.2	–0.1	**–0.8**	–1.0	–2.4	–2.9	–2.8	**–6.6**	–3.7	–2.9	**–3.7**	–3.7
WLSv	–16.5	–11.9	–10.0	–10.6	–10.6	–7.6	–3.4	0.9	1.1	–0.5	–0.5	–1.2	–2.3	–2.9	–3.0	–5.7	–4.3	–3.2	–3.7	–4.2
WLSv-subset	**–16.8**	**–12.1**	–9.8	**–10.8**	**–10.7**	**–7.8**	**–3.5**	1.1	1.2	**–1.0**	**–1.1**	**–1.3**	–2.2	–2.9	**–3.2**	**–6.1**	**–4.4**	–3.0	–3.7	**–4.4**
WLSv-intuitive	**–17.6**	**–12.6**	**–10.1**	–10.5	–10.6	**–8.7**	**–3.8**	**0.7**	1.1	**–0.8**	**–1.9**	**–1.5**	–2.3	**–3.0**	–3.0	**–7.0**	**–4.7**	**–3.3**	–3.7	**–4.3**
WLSv-lasso	**–19.8**	–11.6	–9.7	–10.5	–10.6	**–10.5**	–3.0	1.2	1.2	–0.5	**–1.2**	–1.1	–2.2	–2.9	–3.0	**–7.5**	–4.1	–3.0	–3.7	–4.2
MinT	–25.4	–18.8	–12.4	**–15.3**	**–12.6**	–15.5	–7.0	0.0	–2.0	–2.0	–4.0	–4.6	–4.3	–5.8	–5.1	–11.4	–8.5	–5.0	–7.2	–6.0
MinT-subset	–25.4	–18.8	–12.4	**–15.3**	**–12.6**	–15.5	–7.0	0.0	–2.0	–2.0	–4.0	–4.6	–4.3	–5.8	–5.1	–11.4	–8.5	–5.0	–7.2	–6.0
MinT-intuitive	–25.4	–18.8	–12.4	**–15.3**	**–12.6**	–15.5	–7.0	0.0	–2.0	–2.0	–4.0	–4.6	–4.3	–5.8	–5.1	–11.4	–8.5	–5.0	–7.2	–6.0
MinT-lasso	–25.4	–18.8	–12.4	**–15.3**	**–12.6**	–15.5	–7.0	0.0	–2.0	–2.0	–4.0	–4.6	–4.3	–5.8	–5.1	–11.4	–8.5	–5.0	–7.2	–6.0
MinTs	–25.4	–17.7	–12.1	–14.2	–12.5	–16.1	–6.8	–0.8	–1.6	**–2.4**	–4.0	–4.6	–4.9	–5.9	**–5.2**	–11.6	–8.2	–5.4	–6.8	**–6.2**
MinTs-subset	–25.2	–17.6	–12.1	–14.2	–12.5	–16.1	–6.8	–0.8	–1.6	**–2.4**	–3.9	–4.6	–4.9	–5.9	**–5.2**	–11.5	–8.2	–5.4	–6.8	**–6.2**
MinTs-intuitive	–25.4	–17.7	–12.1	–14.2	–12.5	–16.1	–6.8	–0.8	–1.6	**–2.4**	–4.0	–4.6	–4.9	–5.9	**–5.2**	–11.6	–8.2	–5.4	–6.8	**–6.2**
MinTs-lasso	–25.4	–17.6	–12.1	–14.2	–12.5	–16.1	–6.7	–0.8	–1.6	**–2.4**	–4.0	–4.6	–4.9	–5.9	**–5.2**	–11.6	–8.2	–5.4	–6.8	**–6.2**
EMinT	**–31.2**	**–19.8**	**–12.5**	–14.1	–11.1	**–22.9**	**–10.9**	**–2.4**	**–3.2**	–1.0	**–7.4**	**–7.3**	**–6.9**	**–7.5**	–5.1	**–16.4**	**–11.2**	**–6.9**	**–7.9**	–5.3
Elasso	–31.0	–19.1	–11.1	–13.6	**–11.2**	–22.7	–9.7	–1.8	–2.4	**–1.7**	**–7.4**	–7.2	–6.1	–5.7	–3.5	–16.3	–10.6	–6.0	–6.8	–4.9

Results (Total, A, and BA)

Proportion of time series being selected in Setup 2, with the error correlation being -0.8.

	Top	A	B	AA	AB	BA	BB	Summary
OLS-subset	0.32	0.34	0.95	0.98	1	0.74	1.00	
OLS-intuitive	0.58	0.52	0.93	0.97	1	0.61	0.97	
OLS-lasso	0.61	0.34	0.38	1.00	1	1.00	1.00	
WLSs-subset	0.27	0.40	0.98	1.00	1	0.73	1.00	
WLSs-intuitive	0.49	0.57	0.96	1.00	1	0.74	0.99	
WLSs-lasso	0.48	0.62	0.72	1.00	1	1.00	1.00	
WLSv-subset	0.30	0.42	1.00	1.00	1	0.68	1.00	
WLSv-intuitive	0.49	0.53	0.99	1.00	1	0.47	1.00	
WLSv-lasso	0.35	0.70	0.85	1.00	1	1.00	1.00	
MinT-subset	1.00	1.00	1.00	1.00	1	1.00	1.00	
MinT-intuitive	1.00	1.00	1.00	1.00	1	1.00	1.00	
MinT-lasso	1.00	1.00	1.00	1.00	1	1.00	1.00	
MinTs-subset	0.87	0.85	1.00	1.00	1	0.85	1.00	
MinTs-intuitive	1.00	1.00	1.00	1.00	1	1.00	1.00	
MinTs-lasso	0.86	0.84	1.00	1.00	1	0.85	1.00	
Elasso	0.94	0.79	0.93	1.00	1	1.00	1.00	

Outline

Data description

Australian labour force

- Monthly series from January 2010 to July 2023.
- Hierarchy structure:
  - Top: 1 series
  - State and Territory (STT): 8 series
  - Duration of job search (Duration): 6 series
  - Duration × STT: $n_b = 48$ series
  - $n = 63$ series in total.
- Test set: 2022 Aug-2023 Jul.

Out-of-sample forecast performance (average RMSE)

Out-of-sample forecast results on a single test set (from August 2022 to July 2023).

	Top				Duration				STT				Duration x STT				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	18.5	13.6	18.3	28.3	11.8	12.7	13.9	16.9	6.7	6.0	6.0	6.3	2.3	2.6	2.7	2.9	4.1	4.1	4.4	5.1
BU	**–81.5**	33.4	–19.9	–45.0	**–30.7**	–9.2	–7.9	–10.1	–12.9	–10.4	–13.4	–13.5	0.0	0.0	0.0	0.0	**–17.1**	–2.8	–5.9	–9.3
OLS	–16.2	–14.2	–13.4	–10.4	2.5	–2.6	–2.7	–0.6	–1.8	–0.9	–1.9	0.3	6.7	5.1	5.1	4.9	2.1	0.7	0.4	1.1
OLS-subset	**–17.0**	–2.1	**–31.2**	**–38.4**	**2.0**	–1.7	**–5.0**	**–2.7**	**–2.7**	**–4.2**	**–8.6**	**–7.3**	6.7	5.2	**3.3**	**3.7**	**1.7**	1.1	**–3.5**	**–3.8**
OLS-intuitive	**–79.6**	**–23.9**	**–31.9**	**–32.1**	**–13.0**	0.4	–0.8	0.3	**–8.9**	4.9	7.0	13.2	**6.3**	12.3	12.4	11.6	**–8.5**	5.6	4.7	4.4
OLS-lasso	–16.2	–14.2	–13.4	–10.4	2.5	–2.6	–2.7	–0.6	–1.8	–0.9	–1.9	0.3	6.7	5.1	5.1	4.9	2.1	0.7	0.4	1.1
WLSs	–60.6	–29.4	–44.0	–38.6	–12.0	–7.6	–6.9	–5.9	–6.5	–8.1	–9.4	–8.0	3.2	1.7	1.7	1.6	–7.7	–4.4	–5.8	–5.9
WLSs-subset	**–61.6**	–22.4	**–47.3**	**–50.4**	–12.0	**–8.0**	**–10.5**	**–7.8**	**–6.6**	**–10.7**	**–14.3**	**–12.8**	3.2	5.6	4.1	5.9	**–7.8**	–2.8	**–6.7**	**–6.4**
WLSs-intuitive	–60.6	–29.4	–44.0	–38.6	–12.0	–7.6	–6.9	–5.9	–6.5	–8.1	–9.4	–8.0	3.2	1.7	1.7	1.6	–7.7	–4.4	–5.8	–5.9
WLSs-lasso	–60.6	–29.4	–44.0	–38.6	–12.0	–7.6	–6.9	–5.9	–6.5	–8.1	–9.4	–8.0	3.2	1.7	1.7	1.6	–7.7	–4.4	–5.8	–5.9
WLSv	–60.6	–29.1	–41.4	–36.6	–14.5	–8.7	–5.6	–4.8	–3.3	–6.8	–8.0	–7.0	5.5	2.6	2.6	3.1	–6.7	–4.0	–4.5	–4.6
WLSv-subset	–51.6	**–32.7**	–36.6	–29.6	**–18.3**	**–9.8**	**–10.5**	**–10.9**	–1.1	–4.3	**–8.1**	**–7.3**	**2.5**	**2.1**	**2.3**	**1.8**	**–7.9**	**–4.3**	**–5.8**	**–6.5**
WLSv-intuitive	–60.6	–29.1	–41.4	–36.6	–14.5	–8.7	–5.6	–4.8	–3.3	–6.8	–8.0	–7.0	5.5	2.6	2.6	3.1	–6.7	–4.0	–4.5	–4.6
WLSv-lasso	–60.6	–29.1	–41.4	–36.6	–14.5	–8.7	–5.6	–4.8	–3.3	–6.8	–8.0	–7.0	5.5	2.6	2.6	3.1	–6.7	–4.0	–4.5	–4.6
MinTs	–27.0	–21.1	–22.9	–21.8	–9.1	–7.9	–6.7	–4.6	–3.6	–9.0	–10.5	–7.6	7.7	3.7	3.0	3.4	–1.9	–3.3	–3.9	–3.1
MinTs-subset	**–41.4**	–9.3	–17.8	**–45.1**	**–12.2**	–5.0	**–8.2**	**–6.8**	**–6.1**	**–9.8**	–7.1	**–9.3**	**5.3**	4.7	3.8	3.6	**–5.3**	–1.5	–3.0	**–6.1**
MinTs-intuitive	–27.0	–21.1	–22.9	–21.8	–9.1	–7.9	–6.7	–4.6	–3.6	–9.0	–10.5	–7.6	7.7	3.7	3.0	3.4	–1.9	–3.3	–3.9	–3.1
MinTs-lasso	–27.0	–21.1	–22.9	–21.8	–9.1	–7.9	–6.7	–4.6	–3.6	–9.0	–10.5	–7.6	7.7	3.7	3.0	3.4	–1.9	–3.3	–3.9	–3.1
EMinT	–60.4	–14.0	1.4	–29.9	–6.0	12.0	10.7	–6.7	16.7	–0.9	–12.4	**–21.0**	23.3	17.2	16.7	10.1	7.7	10.8	9.0	–3.7
Elasso	–4.2	–3.3	**–22.3**	–8.0	**–19.7**	**–9.9**	**–19.9**	**–25.3**	**–24.6**	**–24.3**	**–22.6**	–14.6	**–10.8**	**–3.8**	**–0.2**	**–4.9**	**–15.7**	**–9.3**	**–11.4**	**–13.2**

Number of time series being selected

Number of time series being selected.

	Number of time series retained					Optimal parameters		
	Top	Duration	STT	Duration x STT	Total	λ	λ_0	λ_2
None	1	6	8	48	63	-	-	-
OLS-subset	0	5	1	48	54	-	4.16	1.00
WLSs-subset	0	5	1	46	52	-	0.38	0.10
WLSv-subset	1	5	7	48	61	-	0.51	1.00
MinTs-subset	0	1	1	47	49	-	0.03	0.01
Elasso	1	5	2	3	11	213.59	-	-

Outline

1. Hierarchical time series
2. Linear forecast reconciliation
3. Forecast reconciliation with time series selection
4. Simulation experiments
5. Forecasting Australian labour force
6. Conclusions

Conclusions

- Four methods to achieve series selection in forecast reconciliation.
  - **Regularized best-subset selection**
  - Intuitive method
  - Group lasso method
  - **Empirical group lasso method**
- *-Subset and Elasso methods are suggested.
  - Reduce the disparities arising from using different estimates of $\boldsymbol{W}$.
  - Especially effective when dealing with model misspecification issues.
  - When no apparent model misspecification is present, *-Subset and Elasso methods perform well compared to benchmarks.
- Solving L_0-regularized regression problems with additional constraints remains a challenge.

References

- Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Panagiotelis, A. (2023). Forecast reconciliation: A review. International Journal of Forecasting, 40(2), 430-456.
- Ben Taieb, S., & Koo, B. (2019). Regularized regression for hierarchical forecasting without unbiasedness conditions. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1337-1347.
- Hazimeh, H., & Mazumder, R. (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. Operations Research, 68(5), 1517-1537.
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. Computational Statistics & Data Analysis, 55(9), 2579-2589.
- Mazumder, R., Radchenko, P., & Dedieu, A. (2023). Subset selection with shrinkage: Sparse linear modeling when the SNR is low. Operations Research, 71(1), 129-147.
- Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association, 114(526), 804-819.

THAN K YOU

Find me at ...

xqnwang.rbind.io
@Xia0qianWang
@xqnwang
xiaoqian.wang@monash.edu