Cross-temporal probabilistic forecast reconciliation

Daniele Girolimetto
Department of Statistical Sciences, University of Padova
danigiro.github.io
danigiro
daniele.girolimetto@unipd.it

April 16th, 2024

Time series Analysis and Forecasting Society

TAFS

Forthcoming paper

Girolimetto, D., Athanasopoulos, G., Di Fonzo, T., & Hyndman, R. J. (2023). Cross-temporal probabilistic forecast reconciliation: Methodological and practical issues. International Journal of Forecasting. DOI:10.1016/j.ijforecast.2023.10.003

Prof. George Athanasopoulos

Prof. Tommaso Di Fonzo

Prof. Rob J Hyndman

TAFS

Roadmap

1. Introduction
2. Cross-sectional, temporal and cross-temporal framework
3. Point forecast reconciliation (cross-temporal framework)
4. Probabilistic forecast reconciliation (cross-temporal framework)
5. Forecasting the Australian Tourism Demand
6. Forecast reconciliation software
7. How to use forecast reconciliation </>
8. Conclusions

Introduction

- Forecast reconciliation is a post-forecasting process aimed to improve the quality of the base forecasts (however obtained) of a linearly constrained multiple time series by exploiting cross-sectional (e.g., spatial) and/or temporal constraints of the target forecasts

  cross-sectional framework $+$ **temporal** framework $\Rightarrow$ **cross-temporal** framework

- Hot topic on forecasting methodology and practice:

  ➔ several contributions starting from Hyndman *et al.* (2011)

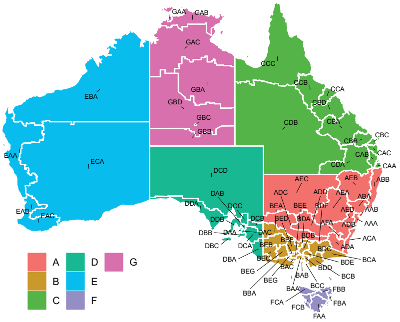
- Many forecasting applications: sales, production, tourism, energy demand, healthcare, real estate, supply chain, macroeconomics, ...

Hands on the Monthly Australian Tourism Demand dataset

- Monthly data on visitor night from 1998 – 2017
- From National Visitor Survey, annual interviews of 120,000 Australians aged 15+
- Source: `robjhyndman.com/data/TourismData_v3.csv`

Cross-sectional framework

Hyndman *et al.* (2011); Panagiotelis *et al.* (2021); Girolimetto and Di Fonzo (2023b)



States and territories of Australia:
7 states and 76 regions

Notation

- $\boldsymbol{y}_t$ = vector of all (n) series at time t
- $\boldsymbol{b}_t$ = vector of the most disaggregated (n_b) series at time t
- $\boldsymbol{a}_t$ = vector of the aggregated $(n_a = n - n_b)$ series at time t
- Two equivalent representations

  1. **Structural representation**: $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$

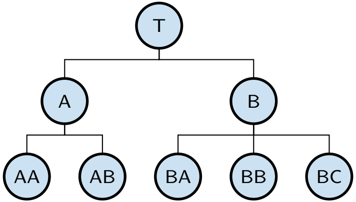
     $\boldsymbol{S} = (n \times n_b)$ "structural matrix"

  2. **Zero-constrained representation**: $\boldsymbol{U}'\mathbf{y}_t = \mathbf{0}_{(n_a \times n)}$

     $\boldsymbol{U}' = (n_a \times n)$ "zero-constrained matrix" containing the linear constraints $(n_a = n - n_b)$

Hierarchical and grouped time series

Genuine hierarchical time series



Constraints

$T = A + B$

$A = AA + AB$

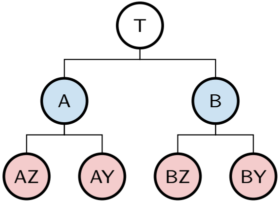
$B = BA + BB + BC$

A cross-sectional hierarchical/grouped time series is a collection of n variables for which - at each time - **aggregation relationships** hold

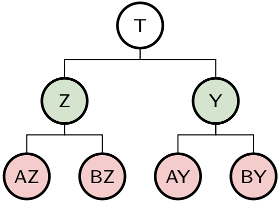
Grouped time series: two or more genuine hierarchies sharing the same top and bottom variables

T	A	B
Z	AZ	BZ
Y	AY	BY

=



+



Constraints

$T = A + B$

$A = AZ + AY$

$B = BZ + BY$

$T = Z + Y$

$Z = AZ + BZ$

$Y = AY + BY$

Grouped time series

Two or more genuine hierarchies sharing the same top and bottom variables

For hierarchical/grouped time series where an aggregation matrix is always available $(\boldsymbol{A})$, the two representations are easily interchangeable:

$$\underbrace{\boldsymbol{C} : \; \boldsymbol{a}_t = \boldsymbol{C}\boldsymbol{b}_t}_{\substack{\text{Linear combination} \\ \text{(or aggregation) matrix}}} \quad \Rightarrow \quad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{I}_{n_b} \end{bmatrix} \quad \text{and} \quad \boldsymbol{U}' = \begin{bmatrix} \boldsymbol{I}_{n_a} & -\boldsymbol{C} \end{bmatrix}$$

Structural and zero-constrained representations in `FoReco`:

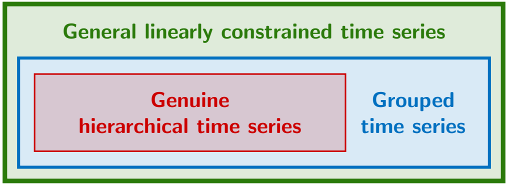
```r
agg_mat <- matrix(c(1,1,1,1,
                    1,1,0,0,
                    0,0,1,1,
                    1,0,1,0,
                    0,1,0,1), nrow = 5, ncol = 4, byrow = TRUE)
hts_tools(C = agg_mat)
```

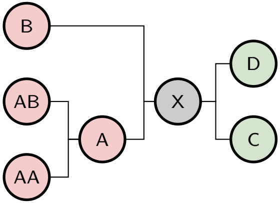
➔

```
List of 6
 |-S : num [1:9, 1:4] 1 1 0 1 0 1 0 0 0 1 ...
 |-C : num [1:5, 1:4] 1 1 0 1 0 1 1 0 0 1 ...
 |-Ut: num [1:5, 1:9] 1 0 0 0 0 0 1 0 0 0 ...
 |-n : int 9
 |-na: int 5
 |-nb: int 4
```

General linearly constrained time series

E.g.: two hierarchies that share only the most most aggregated level





Constraints

$$X = A + B$$
$$X = C + D$$
$$A = AA + AB$$

- Forecast reconciliation may be always expressed according to a **zero-constrained** framework

```r
const_mat <- matrix(c(1, 0, 0,-1,-1, 0, 0,
                      1,-1,-1, 0, 0, 0, 0,
                      0, 1, 0, 0, 0,-1,-1), 3, 7, byrow = TRUE)
hts_tools(Ut = const_mat, nb = 5)
```

→

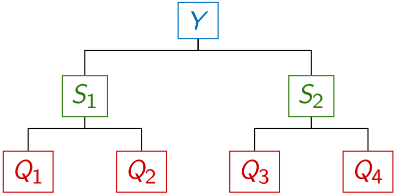
```
List of 6
 |-Ut: num [1:3, 1:7] 1 0 0 0 1 0 0 0 1 -1 ...
 |-n : int 7
 |-na: int 3
 |-nb: int 4
```

- A **structural-like** representation may be derived:

```r
strc_like <- lcmat(const_mat, sparse = FALSE)
```

Temporal framework

Athanasopoulos *et al.* (2017)



Quarterly hierarchy:
quarterly, semi-annual and annual series

Temporal hierarchy ➔ non-overlapping aggregation of the observations of a time series ($\boldsymbol{y}_i$) at regular intervals

$$x_{i,\tau}^{[k]} = \sum_{t=(\tau-1)k+1}^{\tau k} y_{i,t} \qquad \text{for} \qquad \tau = 1, \dots \lfloor T/k \rfloor$$

NB: For $k = 1$, $\tau = t = 1, \dots, T$ and $x_{i,\tau}^{[1]} = y_{i,t}$

- $k \in \mathcal{K} = \{k_p, \dots, k_1\}$ denote the p factors of m in descending order, where $k_1 = 1$ and $k_p = m$
- $k^* = \sum_{k \in \mathcal{K} \setminus \{k_1\}} k$ is the number of upper time series of the temporal hierarchy
- Unlike cross-sectional hierarchies (n **variables at the same time index** are considered), in temporal hierarchies we have **one variable observed at different frequencies**

Temporal matrices: quarterly data

$$\mathbf{x}_{i,\tau} = \begin{bmatrix} x_{i,\tau}^{[m]} \\ \mathbf{x}_{i,\tau}^{[k_{p-1}]} \\ \vdots \\ \mathbf{x}_{i,\tau}^{[k_2]} \\ \mathbf{x}_{i,\tau}^{[1]} = \mathbf{y}_{i,\tau} \end{bmatrix} = \begin{bmatrix} x_{i,\tau}^{[4]} \\ x_{i,2(\tau-1)+1}^{[2]} \\ x_{i,2(\tau-1)+2}^{[2]} \\ y_{i,4(\tau-1)+1} \\ y_{i,4(\tau-1)+2} \\ y_{i,4(\tau-1)+3} \\ y_{i,4(\tau-1)+4} \end{bmatrix}$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{1}'_m \\ \mathbf{I}_{\frac{m}{k_{p-1}}} \otimes \mathbf{1}'_{k_{p-1}} \\ \vdots \\ \mathbf{I}_{\frac{m}{k_2}} \otimes \mathbf{1}'_{k_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\tau = 1, \ldots, \lfloor T/m \rfloor$$

- τ is the time index for the most aggregated series, i is the cross-sectional index, $M_k = m/k$ is the seasonal period of aggregated series
- Structural, $\mathbf{x}_{i,\tau} = \mathbf{R}\mathbf{x}_{i,\tau}^{[1]}$, and zero-constrained, $\mathbf{Z}'\mathbf{x}_{i,\tau} = \mathbf{0}$, representations still hold, and may be alternatively used for reconciliation

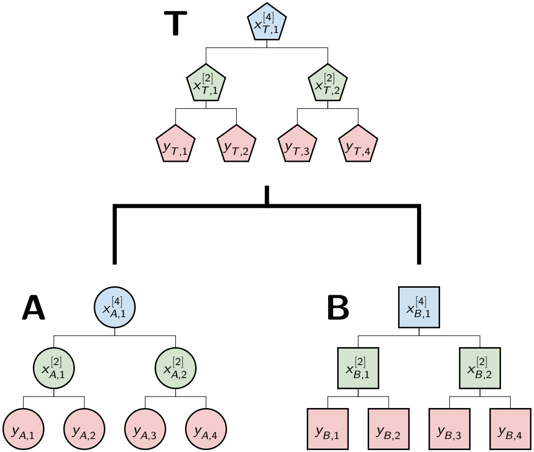
```r
thf_tools(m = 4)
```

```
List of 8
 |-K    : num [1:3, 1:4] 1 1 0 1 1 0 1 0 1 1 ...
 |-Zt   : num [1:3, 1:7] 1 0 0 0 1 0 0 0 1 -1 ...
 |-R    : num [1:7, 1:4] 1 1 0 1 0 0 0 1 1 0 ...
 |-kset: int [1:3] 4 2 1
 |-m    : num 4
 |-p    : int 3
 |-ks   : num 3
 |-kt   : num 7
```

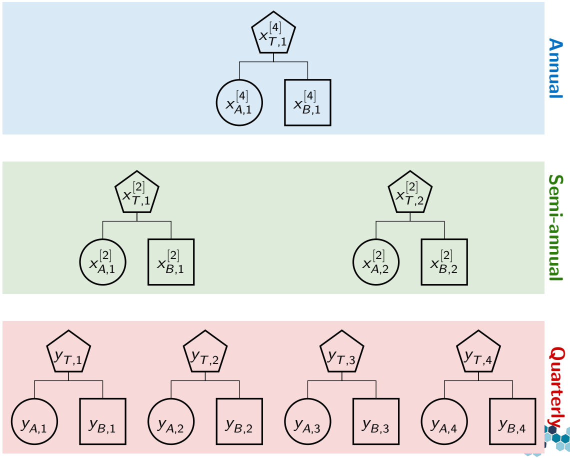
Cross-sectional + Temporal = Cross-temporal

A cross-temporal hierarchy of three quarterly time series ($T = A + B$)

cross-sectional $\longrightarrow$ temporal



temporal $\longrightarrow$ cross-sectional



Cross-temporal framework

Di Fonzo and Girolimetto (2023a)

$$
\boldsymbol{X}_\tau = \begin{bmatrix} \boldsymbol{x}_{1,\tau}' \\ \vdots \\ \boldsymbol{x}_{n,\tau}' \end{bmatrix} = \begin{array}{c}
\begin{array}{ccc} \color{blue}{\boldsymbol{U}^{[4]}} & \color{green}{\boldsymbol{U}^{[2]}} & \color{red}{\boldsymbol{U}^{[1]}} \end{array} \\
\left[\begin{array}{c|cc|cccc}
x_{T,\tau}^{[4]} & x_{T,2\tau-1}^{[2]} & x_{T,2\tau}^{[2]} & y_{T,4\tau-3} & y_{T,4\tau-2} & y_{T,4\tau-1} & y_{T,4\tau} \\
x_{A,\tau}^{[4]} & x_{A,2\tau-1}^{[2]} & x_{A,2\tau}^{[2]} & y_{A,4\tau-3} & y_{A,4\tau-2} & y_{A,4\tau-1} & y_{A,4\tau} \\
x_{B,\tau}^{[4]} & x_{B,2\tau-1}^{[2]} & x_{B,2\tau}^{[2]} & y_{B,4\tau-3} & y_{B,4\tau-2} & y_{B,4\tau-1} & y_{B,4\tau}
\end{array}\right] \\
\begin{array}{ccc} \color{blue}{\boldsymbol{B}^{[4]}} & \color{green}{\boldsymbol{B}^{[2]}} & \color{red}{\boldsymbol{B}^{[1]}} \end{array}
\end{array}
$$

$$
\boldsymbol{x}_\tau = \text{vec}(\boldsymbol{X}_\tau') \qquad \boldsymbol{b}_\tau^{[1]} = \text{vec}(\boldsymbol{B}_\tau^{[1]\prime})
$$

Any cross-temporal matrix may be constructed starting from the one-dimensional equivalents

$\boldsymbol{H}' \rightarrow$ easy to compute as a function of $\boldsymbol{U}'$ and of the highest time frequency (m)

$$
\boldsymbol{U}'\boldsymbol{X}_\tau\boldsymbol{Z} = \boldsymbol{0}_{(n_a \times k^*)} \leftrightarrow \boldsymbol{H}'\boldsymbol{x}_\tau = \boldsymbol{0}_{[(n_a m + nk^*)\times 1]}
$$

$\boldsymbol{F} \rightarrow$ fast to compute as $\boldsymbol{S} \otimes \boldsymbol{R}$

$$
\boldsymbol{X}_\tau = \boldsymbol{S}\boldsymbol{B}_\tau^{[1]}\boldsymbol{R}' \leftrightarrow \boldsymbol{x}_\tau = \boldsymbol{F}\boldsymbol{b}_\tau^{[1]}
$$

```r
agg_mat <- matrix(c(1,1), nrow = 1, ncol = 1)
ctf_tools(C = agg_mat, m = 4)
```

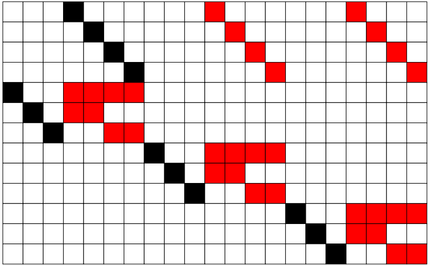
```
List of 3
 |-ctf:List of 6
 |  |-Ht      : num [1:36, 1:56] 0 0 0 0 0 0 0 0 0 0 ...
 |  |-Hbrevet: num [1:45, 1:56] 1 0 0 0 0 0 0 0 0 0 ...
 |  |-Hcheckt: num [1:36, 1:56] 1 0 0 0 0 0 0 0 0 0 ...
 |  |-Ccheck : num [1:36, 1:20] 1 1 0 1 0 0 0 1 1 0 ...
 |  |-Scheck : num [1:56, 1:20] 1 1 0 1 0 0 0 1 1 0 ...
 |  |-Fmat   : num [1:56, 1:20] 1 1 0 1 0 0 0 1 1 0 ...
 |-hts:List of 6
 |-thf:List of 8
```

Cross-temporal representations

Two dimensions (**spatio–temporal**) to capture the complete nature of a multiple time series

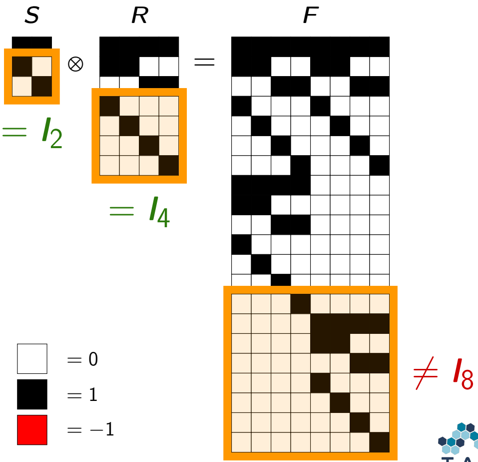
Zero-constrained representation

$$\boldsymbol{H}' = \begin{bmatrix} (\boldsymbol{0}_{(n_a m \times nk^*)} \quad \boldsymbol{I}_m \otimes \boldsymbol{U}')\boldsymbol{P}' \\ \boldsymbol{I}_n \otimes \boldsymbol{Z}' \end{bmatrix}$$



where $\quad \boldsymbol{P}\mathrm{vec}(\boldsymbol{X}_\tau) = \mathrm{vec}(\boldsymbol{X}'_\tau)$

Structural representation



Point forecast reconciliation

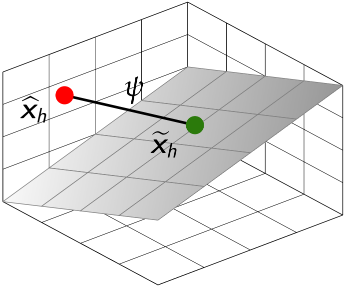
Wickramasuriya *et al.* (2019); Panagiotelis *et al.* (2021)

Definition

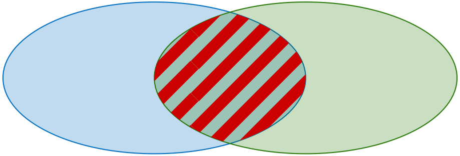
Forecast reconciliation aims to adjust the base forecast $\widehat{\boldsymbol{x}}_h$ via a mapping $\psi : \mathbb{R}^{n(m+k^*)} \rightarrow \mathfrak{s}$ such that

$$\widetilde{\boldsymbol{x}}_h = \psi\left(\widehat{\boldsymbol{x}}_h\right),$$

where $\widetilde{\boldsymbol{x}}_h \in \mathfrak{s}$ is the vector of the reconciled forecasts



$\mathfrak{S}_{cs}$
cross-sectional *coherent subspace* spanned by the columns of $\boldsymbol{S}$



$\mathfrak{S}_{te}$
temporal *coherent subspace* spanned by the columns of $\boldsymbol{R}$

$\mathfrak{S}$ cross-temporal *coherent subspace* spanned by the columns of $\boldsymbol{F} = \boldsymbol{S} \otimes \boldsymbol{R}$

Optimal forecast reconciliation

Base forecasts	Target		Reconciled forecasts
$\boldsymbol{H}'\widehat{\boldsymbol{x}}_h \neq \boldsymbol{0}$	$\boldsymbol{H}'\boldsymbol{x}_h = \boldsymbol{0}$	$\rightarrow$	$\boldsymbol{H}'\widetilde{\boldsymbol{x}}_h = \boldsymbol{0}$

1. Forecast **all series at all levels** of aggregation (using e.g. ARIMA, ETS, VAR ...) $\rightarrow$ base forecasts
2. Make the base forecasts **coherent** $\rightarrow$ reconciled forecasts
   - **Projection approach** (Byron, 1978, 1979)

$$\begin{aligned} \arg\min_{\boldsymbol{x}_h} \quad & (\widehat{\boldsymbol{x}}_h - \boldsymbol{x}_h)'\boldsymbol{\Omega}_{ct}^{-1}(\widehat{\boldsymbol{x}}_h - \boldsymbol{x}_h) \\ \text{s.t.} \quad & \boldsymbol{H}'\boldsymbol{x}_h = \boldsymbol{0} \end{aligned} \quad \Rightarrow \quad \widetilde{\boldsymbol{x}}_h = \left[\boldsymbol{I} - \boldsymbol{\Omega}_{ct}\boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{\Omega}_{ct}\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right]\widehat{\boldsymbol{x}}_h = \psi\left(\widehat{\boldsymbol{x}}_h\right)$$

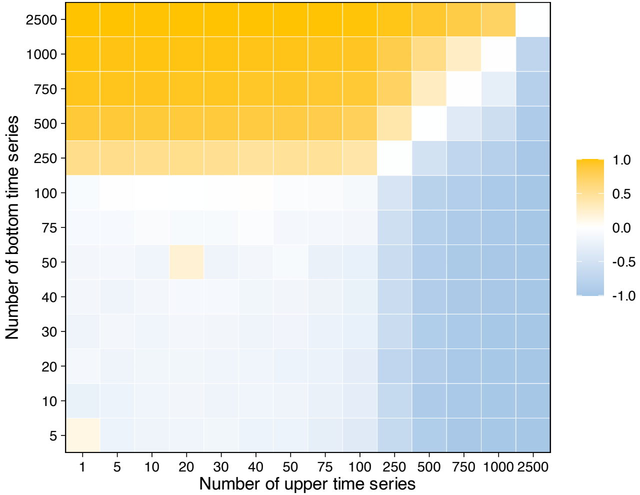
   - **Structural approach** (Wickramasuriya *et al.*, 2019, cross-temporal extension)

$$\begin{aligned} \min_{\boldsymbol{G}} \quad & \operatorname{tr}\left(\boldsymbol{F}\boldsymbol{G}\boldsymbol{\Omega}_{ct}\boldsymbol{G}'\boldsymbol{S}_{ct}'\right) \\ \text{s.t.} \quad & \boldsymbol{F}\boldsymbol{G}\boldsymbol{F} = \boldsymbol{F} \end{aligned} \quad \Rightarrow \quad \widetilde{\boldsymbol{x}}_h = \boldsymbol{F}\underbrace{\left(\boldsymbol{F}'\boldsymbol{\Omega}_{ct}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\boldsymbol{\Omega}_{ct}^{-1}}_{\boldsymbol{G}}\widehat{\boldsymbol{x}}_h = \psi\left(\widehat{\boldsymbol{x}}_h\right)$$

The formulation of $\text{Var}(\widehat{\boldsymbol{x}}_h - \boldsymbol{x}_h)$ is conceptually complex; in practice, approximate forms of $\text{Var}(\widehat{\boldsymbol{x}}_h - \boldsymbol{x}_h) \approx \boldsymbol{\Omega}_{ct}$ are used, possibly using in-sample residuals

Projection vs structural approach

FoReco 0.2.6 - cross-sectional ols reconciliation - *time* in seconds



Two main factors:

- dimensions

  cs: $n_a,\ n_b$

  te: $\mathcal{K}$

  ct: $n_a,\ n_b,\ \mathcal{K}$

- Computational cost of $\boldsymbol{\Omega}_{ct}^{-1}$

$$value = \frac{time_{\text{strc}} - time_{\text{proj}}}{time_{\text{strc}} + time_{\text{proj}}}$$

$value < 0 \Rightarrow$ structural is faster

$value > 0 \Rightarrow$ projection is faster

Probabilistic forecast reconciliation

Starting point

Panagiotelis *et al.* (2023) definitions, theorems and approaches for the cross-sectional case $\rightsquigarrow$ Coherence and probabilistic forecast reconciliation according to the more general cross-temporal framework

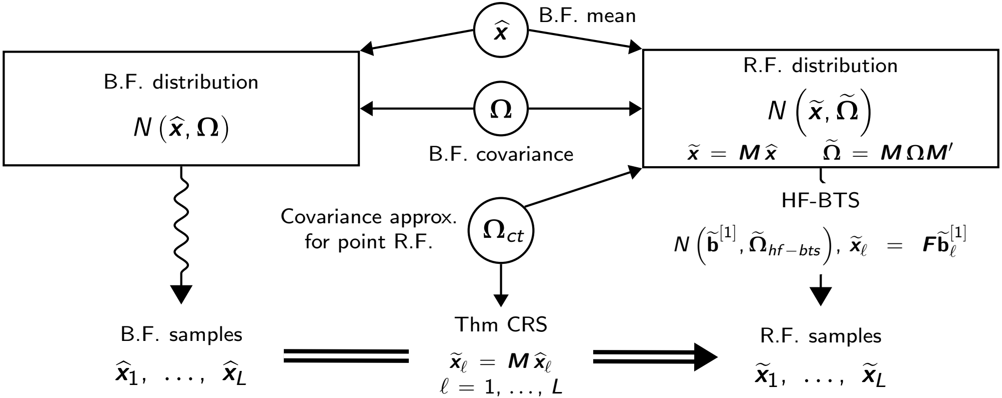
- A generalized probabilistic cross-temporal framework for **count data** can be obtained following the cross-sectional work of Corani *et al.* (2024) $\rightarrow$ however, we only focus on the **continuous case**

Thm: Cross-temporal Reconciled Samples

$(\widehat{\boldsymbol{x}}_1, \dots, \widehat{\boldsymbol{x}}_L)$ sample drawn from an incoherent probability measure $\widehat{\nu}$ $\xrightarrow{\widetilde{\boldsymbol{x}}^{[\ell]} = \psi\left(\widehat{\boldsymbol{x}}^{[\ell]}\right)}$ $(\widetilde{\boldsymbol{x}}_1, \dots, \widetilde{\boldsymbol{x}}_L)$ sample drawn from the reconciled probability measure $\widetilde{\nu}$

A parametric gaussian approach

Base forecasts sample paths (Panagiotelis *et al.*, 2023; Wickramasuriya, 2024)



- $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_{ct}$ estimates may consider cross-sectional and/or temporal dimensions (G, H, B, HB)
- Using in-sample residuals is challenging ➔ **multi-step residuals**

A non-parametric bootstrap approach

Base forecasts sample paths (Panagiotelis *et al.*, 2023)

- Analytical expressions for the base and reconciled forecast distributions are challenging and parametric assumptions can be restrictive and unrealistic
- **Joint (block) Bootstrap:** simulate future base sample paths from all models using bootstrapped residuals, then **reconcile** them to obtain coherent sample paths

$$\widehat{\boldsymbol{x}}_{i,\ell}^{[k]} = f_i(\mathcal{M}_i, \widehat{\boldsymbol{e}}_{i,\tau}^{[k]})$$

- Need to generate samples that preserve **cross-temporal** relationships

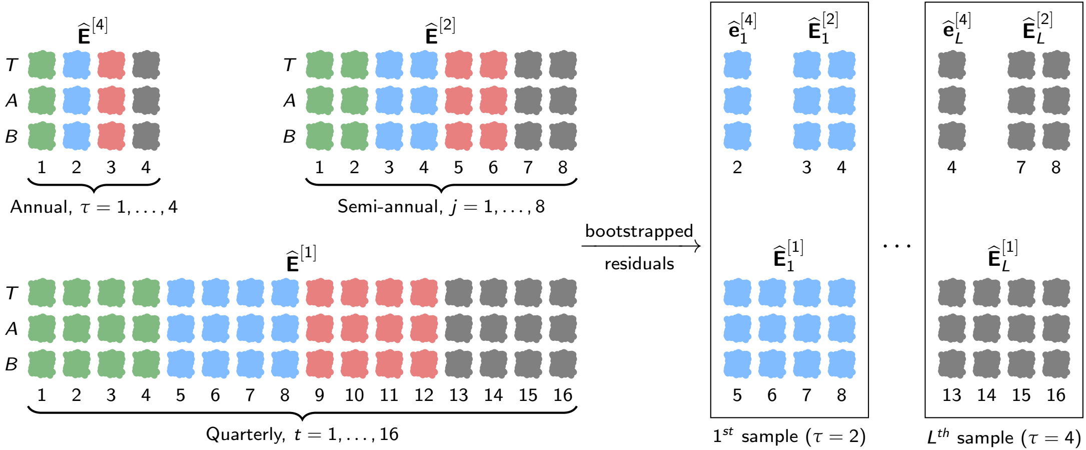
Idea

1. sampling the index time τ from the most temporally aggregated level
2. using it to determine the **indices for all the other levels**

+ Easily scalable in order to utilize multiple computing power simultaneously for each series
+ Take into account the dependence between the different levels of temporal aggregation and not only the cross-sectional dependencies

Example of bootstrapped residuals

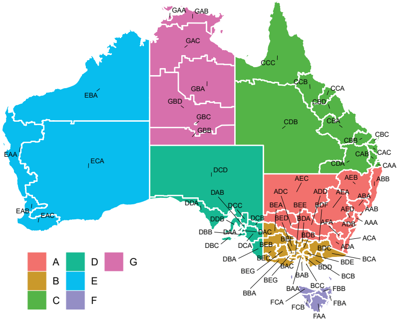
$T = A + B$, 16 quarterly data: the green, blue, red and black colors correspond, respectively, to years 1, 2, 3 and 4.



Forecasting the Australian Tourism Demand

Classical dataset in the hierarchical forecasting literature

Geographical division



✖

Purpose of travel

Holiday, Visiting friends and relatives, Business, Other

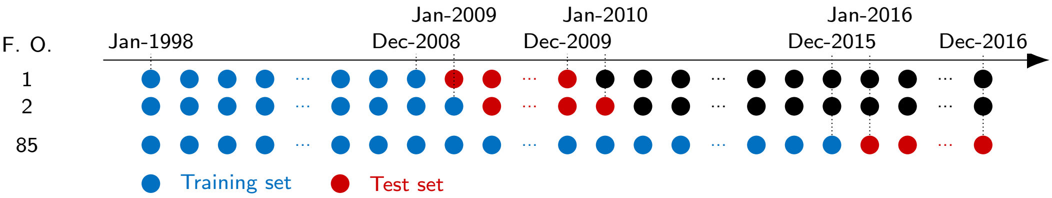
- **Grouped ts** (geographical divisions $\times$ purpose of travel)

	AUS	States	Zones*	Regions	Tot
g.d.	1	7	21	76	105
p.o.t.	4	28	84	304	420
Tot	5	35	105	380	**525**

$n_a = 221$, $n_b = 304$, and $\mathbf{n} = \mathbf{525}$

- **Unique time series, no redundancy** (*6 Zones with only one Region are included in the Regions)
- **Temporal framework**, frequencies:
  - Monthly
  - Bi-Monthly
  - Quarterly
  - Four-Monthly
  - Semi-Annual
  - Annual

The forecasting experiment



- Monthly data: expanding window, monthly step and 12-step ahead forecast horizons ($h_1 = 12$)
- For each training set, temporally aggregated series for any $k \in \mathcal{K}$ are computed, and forecasts are produced up to $h_2 = 6$, $h_3 = 4$, $h_4 = 3$, $h_6 = 2$ and $h_{12} = 1$ step ahead, respectively
- Automatic ETS forecasts on log-transformed data (Wickramasuriya *et al.*, 2020)
- **Accuracy indices** (Gneiting and Katzfuss, 2014)
  - Continuous Ranked Probability Score (CRPS)
  - Energy Score (ES)

Dealing with **negativity issues**: set-negative-to-zero (Di Fonzo and Girolimetto, 2023b)

Probabilistic forecasts sample paths

All the reconciliation procedures are available in `FoReco`

Base forecasts sample paths

- Gaussian approach (4 variants)
- Cross-temporal Joint (block) Bootstrap (ctjb)

Reconciliation approaches

- Cross-temporal **bottom-up** and **partly bottom-up**

  ct(bu) | ct(shr_{cs}, bu_{te}) | ct$(wlsv_{te}, bu_{cs})$

- Optimal forecast reconciliation with **one-step residuals** (Di Fonzo and Girolimetto, 2023a)

  oct(ols) | oct$(struc)$ | oct$(wlsv)$ | oct$(bdshr)$

TAFS

RelCRPS for the Australian Tourism Demand dataset

Red: worse than the benchmark (ctjb). Bold: the best for each column. Blue: the overall lowest value

Reconciliation approach	ctjb	Gaussian approach G	Gaussian approach B	Gaussian approach H	Gaussian approach HB	ctjb	Gaussian approach G	Gaussian approach B	Gaussian approach H	Gaussian approach HB
	All temporal level, $\forall k \in \{12, 6, 4, 3, 2, 1\}$					**Monthly level, $k = 1$**				
base	1.000	0.971	0.971	0.973	0.973	1.000	0.972	0.972	0.972	0.972
ct(bu)	1.321	1.011	1.011	1.011	1.011	1.077	0.983	0.982	0.982	0.982
ct(shr_{cs}, bu_{te})	1.057	0.974	0.969	0.974	0.969	0.976	0.963	0.962	0.963	0.962
ct($wlsv_{te}$, bu_{cs})	1.062	0.974	0.974	0.972	0.972	0.976	0.965	0.965	0.966	0.966
oct(ols)	0.989	0.989	0.989	0.987	0.987	0.982	0.986	0.988	0.986	0.989
oct($struc$)	0.982	0.962	0.961	0.961	0.959	0.970	0.963	0.963	0.963	0.963
oct($wlsv$)	0.987	0.959	0.959	0.958	0.957	0.952	0.957	0.957	0.957	0.957
oct($bdshr$)	**0.975**	**0.956**	**0.953**	**0.952**	**0.951**	**0.949**	**0.955**	**0.953**	**0.954**	**0.954**

Overall, oct($bdshr$) in terms of CRPS is almost always the best → MCB Nemenyi test

Forecast reconciliation software

Source: "Probabilistic cross-temporal forecast reconciliation" by Prof. Rob J Hyndman at the *International Association of Statistical Computing: Asian Regional Section Conference 2023*

Package	Language	Cross-sectional	Temporal	Cross-temporal	Probabilistic
`hts`	R	✔			
`thief`	R		✔		
`fable`*	R	✔			✔
`FoReco`	R	✔	✔	✔	✔
`pyhts`	Python	✔	✔		
`hierarchicalforecast`	Python	✔			✔

* `fable` has plans to implement temporal and cross-temporal reconciliation

What is FoReco?

R package (Girolimetto and Di Fonzo, 2023a)

- FoReco offers classical (bottom-up and top-down), and modern (optimal and heuristic combination) **forecast reconciliation procedures** for cross-sectional, temporal, and **cross-temporal** **linearly constrained multiple time series**.
- Matrix-based package, exploiting the very sparse nature of the involved matrices
- **Links**:
  - cran.r-project.org/package=FoReco
  - github.com/daniGiro/FoReco
  - danigiro.github.io/FoReco

FoReco

Available on R

First release: 01/10/2020
Last release: 16/05/2023

New release? Soon!

TAFS

How to use forecast reconciliation </>

Daniele Girolimetto
Department of Statistical Sciences, University of Padova
`danigiro.github.io`
`danigiro`
@ `daniele.girolimetto@unipd.it`

April 16th, 2024

Time series Analysis and Forecasting Society

Resources about forecast reconciliation

Seminars presented by by Prof. Rob J Hyndman - `robjhyndman.com`

Rob J Hyndman Hyndsight blog Publications Software Seminars Teaching Research Team About

On this page

Seminars I've given (since 2006)

Seminars I've given (since 2006)

Some of these seminars were given more than once, but to avoid repetition, repeats are not always listed.

Talks on YouTube are available on this playlist.

Filter

Date	Title	Venue
08 Apr 2024	Developing good research habits	Monash
08 Feb 2024	Forecasting the future and the future of forecasting	Victorian Treasury Theatre
14 Dec 2023	Probabilistic Forecast Reconciliation For Emergency Services Demand	Australian Statistics Conference
08 Dec 2023	Probabilistic cross-temporal forecast reconciliation	International Association of Statistical Computing: Asian Regional Section Conference 2023
07 Nov 2023	Forecast reconciliation	Online
11 Sep 2023	Forecast reconciliation: a brief overview	Zalando, Berlin, Germany

TAFS

Awesome Forecast Reconciliation - Github repository

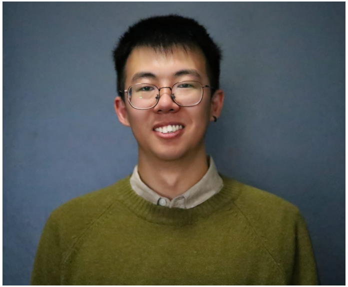
Awesome Forecast Reconciliation

This repository serves as a curated reference for the domain of forecast reconciliation. It aims to contain an extensive collection of academic papers, articles, software tools, and educational resources. Ideal for researchers, analysts, and practitioners seeking to improve the consistency and precision of forecasting methodologies.

We wish to express our deep appreciation to the authors of the paper "Forecast reconciliation: A review" - George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Anastasios Panagiotelis - for providing their BibTeX file, which served as the cornerstone of this repository. Their paper serves as an invaluable resource with its comprehensive and insightful analysis of the forecast reconciliation field, providing a thorough overview of the existing literature and highlighting key advancements and research trends.

⚠️ The list is still incomplete and unorganized. We are in the process of reorganizing the various items.

`danigiro/awesome-forecast-reconciliation`



Yangzhuoran Fin Yang
Monash University

April 16th 2024

Time series Analysis and Forecasting Society

UNIVERSITÀ DEGLI STUDI DI PADOVA

THANK YOU!

Daniele Girolimetto

Department of Statistical Sciences

University of Padova (Italy)

danigiro

danigiro.github.io

@ daniele.girolimetto@unipd.it

FoReco

Awesome FR

References I

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N. and Petropoulos, F. (2017) Forecasting with temporal hierarchies. *European Journal of Operational Research* **262**(1), 60–74

Byron, R. P. (1978) The Estimation of Large Social Account Matrices. *Journal of the Royal Statistical Society. Series A (General)* **141**(3), 359–367

Byron, R. P. (1979) Corrigenda: The Estimation of Large Social Account Matrices. *Journal of the Royal Statistical Society. Series A (General)* **142**(3), 405

Corani, G., Azzimonti, D. and Rubattu, N. (2024) Probabilistic reconciliation of count time series. *International Journal of Forecasting* **40**(2), 457–469

Di Fonzo, T. and Girolimetto, D. (2022) Enhancements in cross-temporal forecast reconciliation, with an application to solar irradiance forecasts. *arXiv:2209.07146*

Di Fonzo, T. and Girolimetto, D. (2023a) Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. *International Journal of Forecasting* **39**(1), 39–57

Di Fonzo, T. and Girolimetto, D. (2023b) Spatio-temporal reconciliation of solar forecasts. *Solar Energy* **251**, 13–29

Girolimetto, D. and Di Fonzo, T. (2023a) *FoReco: Point Forecast Reconciliation*. R package v0.2.6 **URL:** `https://danigiro.github.io/FoReco/`

Girolimetto, D. and Di Fonzo, T. (2023b) Point and probabilistic forecast reconciliation for general linearly constrained multiple time series. *Statistical Methods & Applications* (in press)

Gneiting, T. and Katzfuss, M. (2014) Probabilistic Forecasting. *Annual Review of Statistics and Its Application* **1**(1), 125–151

References II

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. and Shang, H. L. (2011) Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* **55**(9), 2579–2589

Ledoit, O. and Wolf, M. (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**(2), 365–411

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P. and Hyndman, R. J. (2021) Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting* **37**(1), 343–359

Panagiotelis, A., Gamakumara, P., Athanasopoulos, G. and Hyndman, R. J. (2023) Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. *European Journal of Operational Research* **306**(2), 693–706

Stellato, B., Banjac, G., Goulart, P., Bemporad, A. and Boyd, S. (2020) OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation* **12**(4), 637–672

Wickramasuriya, S. L. (2024) Probabilistic forecast reconciliation under the gaussian framework. *Journal of Business & Economic Statistics* **42**(1), 272–285

Wickramasuriya, S. L., Athanasopoulos, G. and 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

Wickramasuriya, S. L., Turlach, B. A. and Hyndman, R. J. (2020) Optimal non-negative forecast reconciliation. *Statistics and Computing* **30**(5), 1167–1182

Projection approach

Di Fonzo and Girolimetto (2023a)

- **Multivariate constrained regression model**

$$\begin{gathered}\widehat{\boldsymbol{X}}_h = \boldsymbol{X}_h + \boldsymbol{E}_h \\ \text{s.t. } \boldsymbol{U}'\boldsymbol{X}_h = \boldsymbol{0}_{(n_a^* \times 1)} \text{ and } \boldsymbol{Z}'\boldsymbol{X}_h' = \boldsymbol{0}_{(k^* \times 1)}\end{gathered} \iff \widehat{\boldsymbol{x}}_h = \boldsymbol{x}_h + \boldsymbol{\eta}_h \quad \text{s.t.} \quad \boldsymbol{H}'\boldsymbol{x}_h = \boldsymbol{0}_{[(n_a m + nk^*) \times 1]}$$

where $\widehat{\boldsymbol{x}}_h = \text{vec}\left(\widehat{\boldsymbol{X}}_h'\right)$, $\boldsymbol{x}_h = \text{vec}\left(\boldsymbol{X}_h'\right)$ and $\boldsymbol{\eta}_h = \text{vec}\left(\boldsymbol{E}_h'\right)$ such that $E\left[\boldsymbol{\eta}_h\right] = \boldsymbol{0}_{[n(m+k^*) \times 1]}$ and $\text{Var}\left(\boldsymbol{\eta}_h\right) = \boldsymbol{\Omega}_{ct}$

- **Solution** (Byron, 1978, 1979, cross-temporal extension)

$$\min_{\boldsymbol{y}_h} \; (\widehat{\boldsymbol{x}}_h - \boldsymbol{x}_h)'\boldsymbol{\Omega}_{ct}^{-1}(\widehat{\boldsymbol{x}}_h - \boldsymbol{x}_h) \quad \text{s.t.} \quad \boldsymbol{H}'\boldsymbol{x}_h = \boldsymbol{0}_{[(n_a m + nk^*) \times 1]}$$

$$\Rightarrow \quad \widetilde{\boldsymbol{x}}_h = \left[\boldsymbol{I} - \boldsymbol{\Omega}_{ct}\boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{\Omega}_{ct}\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right]\widehat{\boldsymbol{x}}_h = \psi\left(\widehat{\boldsymbol{x}}_h\right)$$

TAFS

Structural approach

Hyndman *et al.* (2011); Wickramasuriya *et al.* (2019); Di Fonzo and Girolimetto (2023a)

- **Multivariate regression model**

$$\widehat{\boldsymbol{X}}_h = \boldsymbol{S}\boldsymbol{B}_h^{[1]}\boldsymbol{R}' + \boldsymbol{E}_h \quad \Longleftrightarrow \quad \widehat{\boldsymbol{x}}_h = \boldsymbol{F}\boldsymbol{b}_h^{[1]} + \boldsymbol{\eta}_h$$

where $\widehat{\boldsymbol{x}}_h = \text{vec}\left(\widehat{\boldsymbol{X}}_h'\right)$, $\boldsymbol{b}_h^{[1]} = \text{vec}\left(\boldsymbol{B}_h^{[1]'}\right)$ and $\boldsymbol{\eta}_h = \text{vec}\left(\boldsymbol{E}_h'\right)$ such that $E\left[\boldsymbol{\eta}_h\right] = \boldsymbol{0}_{[n(m+k^*)\times 1]}$ and $\text{Var}\left(\boldsymbol{\eta}_h\right) = \boldsymbol{\Omega}_{ct}$

- **Solution** (Wickramasuriya *et al.*, 2019, cross-temporal extension)

$$\min_{\boldsymbol{G}} \ \text{tr}\left(\boldsymbol{F}\boldsymbol{G}\boldsymbol{\Omega}_{ct}\boldsymbol{G}'\boldsymbol{S}_{ct}'\right) \qquad \text{s.t.} \quad \boldsymbol{F}\boldsymbol{G}\boldsymbol{F} = \boldsymbol{F}$$

$$\Rightarrow \quad \widetilde{\boldsymbol{x}}_h = \boldsymbol{F}\underbrace{\left(\boldsymbol{F}'\boldsymbol{\Omega}_{ct}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\boldsymbol{\Omega}_{ct}^{-1}}_{\boldsymbol{G}}\widehat{\boldsymbol{x}}_h = \psi\left(\widehat{\boldsymbol{x}}_h\right)$$

with $\text{Var}(\boldsymbol{x}_h - \widetilde{\boldsymbol{x}}_h) = \boldsymbol{F}\boldsymbol{G}\boldsymbol{\Omega}_{ct}\boldsymbol{G}'\boldsymbol{S}_{ct}'$ if $\widetilde{\boldsymbol{x}}_h = \boldsymbol{F}\boldsymbol{G}\widehat{\boldsymbol{x}}_h$

Accuracy indices for probabilistic forecasts

Gneiting and Katzfuss (2014)

Continuous Ranked Probability Score

$$\mathrm{CRPS}(\widehat{P}_i, z_i) = \frac{1}{L}\sum_{l=1}^{L} |x_{i,l} - z_i| - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{j=1}^{L} |x_{i,l} - x_{i,j}|$$

Energy Score

$$\mathrm{ES}(\widehat{P}, \boldsymbol{z}) = \frac{1}{L}\sum_{l=1}^{L} \|\boldsymbol{x}_l - \boldsymbol{z}\|_2 - \frac{1}{2(L-1)}\sum_{i=1}^{L-1} \|\boldsymbol{x}_l - \boldsymbol{x}_{l+1}\|_2$$

- $\widehat{P}_i(\omega) = \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}\left(x_{i,l} \leq \omega\right)$
- $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L$ is a collection of L random draws taken from the predictive distribution
- $\boldsymbol{z} \in \mathbb{R}^n$ is the observation vector
- $i = 1, \ldots, n$ denotes a single variable

Cross-temporal covariance matrix estimation

- As $\boldsymbol{\Omega}_{ct}$ (and $\boldsymbol{\Omega}$) is unknown in practice $\rightarrow$ empirical sample covariance of the base forecasts $\widehat{\boldsymbol{\Omega}}$

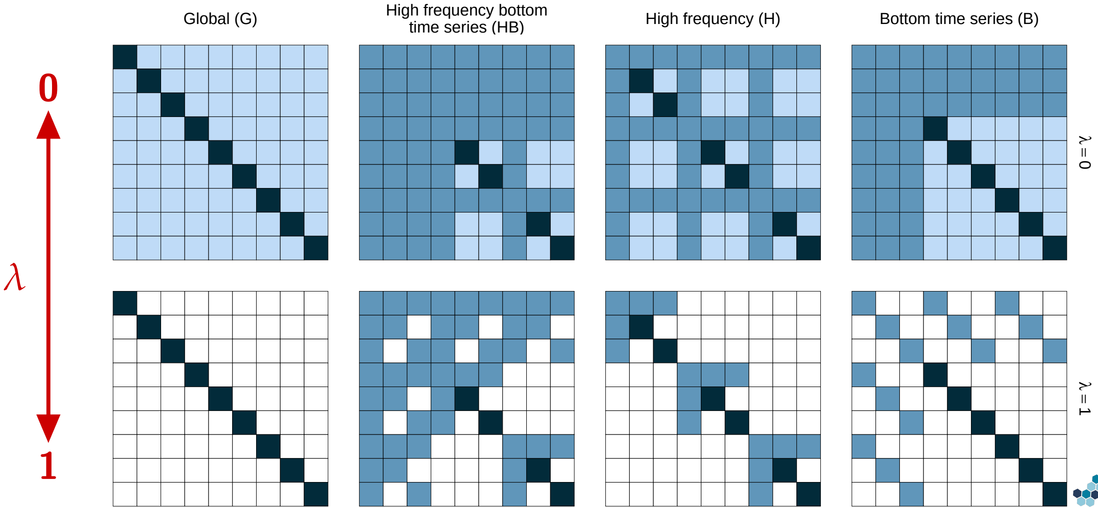
$$\frac{n(k^* + m)\left[n(k^* + m) - 1\right]}{2} \quad \text{different parameters}$$

- A possible solution to estimating many parameters, is to construct a shrinkage estimator using a convex combination of $\widehat{\boldsymbol{\Omega}}$ and a diagonal target matrix $\widehat{\boldsymbol{\Omega}}_D = \widehat{\boldsymbol{\Omega}} \odot \boldsymbol{I}_{n(k^*+m)}$
- Shrinking all off-diagonal elements to zero, when we know that the covariance matrix has a cross-sectional and/or temporal structure, results in information loss
- Use the cross-sectional and/or temporal structure to obtain a better estimator for the covariance matrix of the entire system

$$\widetilde{\boldsymbol{\Omega}}_{ct} = \underbrace{\boldsymbol{F}\boldsymbol{\Omega}_{hf\text{-}bts}\boldsymbol{F}'}_{\text{Cross-temporal structure}} = \underbrace{(\boldsymbol{I}_n \otimes \boldsymbol{R})\,\boldsymbol{\Omega}_{hf}\,(\boldsymbol{I}_n \otimes \boldsymbol{R})'}_{\text{Temporal structure}} = \underbrace{(\boldsymbol{S} \otimes \boldsymbol{I}_{m+k^*})\,\boldsymbol{\Omega}_{bts}\,(\boldsymbol{S} \otimes \boldsymbol{I}_{m+k^*})'}_{\text{Cross-sectional structure}}$$

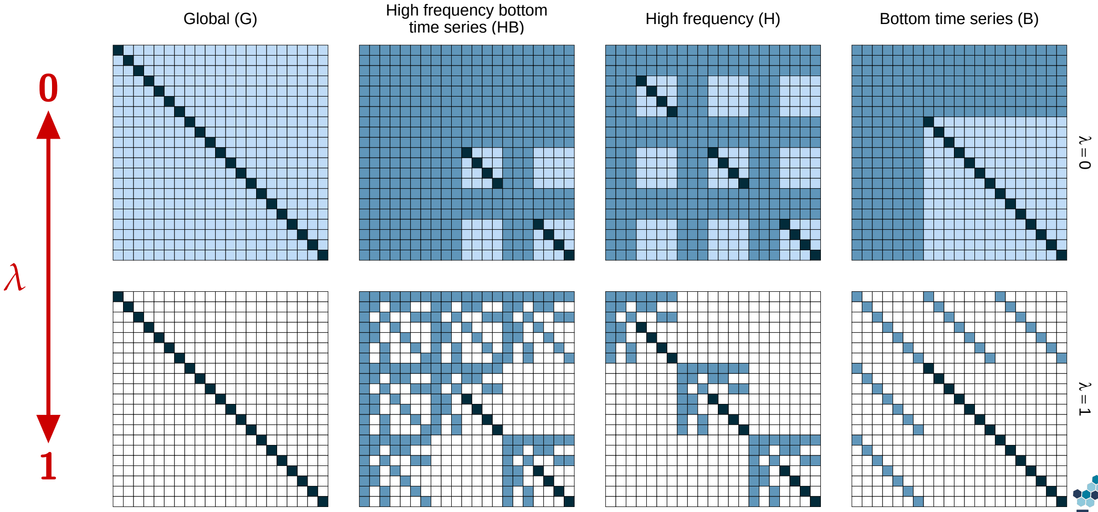
Shrinking covariance matrix representation

Hierarchical structure with 3 time series and $m = 4$ with two different values of $\lambda \in \{0, 1\}$, the shrinkage parameter



Shrinking covariance matrix representation

Hierarchical structure with 3 time series and $m = 2$ with two different values of $\lambda \in \{0, 1\}$, the shrinkage parameter



ES ratio indices for the Australian Tourism Demand dataset

Red: worse than the benchmark (ctjb). Bold: the best for each column. Blue: the overall lowest value.

Reconciliation approach	Generation of the base forecasts sample paths									
	ctjb	Gaussian approach				ctjb	Gaussian approach			
		G	B	H	HB		G	B	H	HB
	$\forall k \in \{12, 6, 4, 3, 2, 1\}$					$k = 1$				
base	1.000	0.956	0.955	0.958	0.951	1.000	0.952	0.950	0.952	0.950
ct(bu)	2.427	0.983	0.983	0.983	0.983	1.759	0.982	0.982	0.982	0.982
ct(shr_{cs}, bu_{te})	1.243	0.886	0.879	0.886	0.879	1.098	0.929	0.928	0.930	0.927
ct($wlsv_{te}$, bu_{cs})	1.499	0.977	0.977	0.971	0.972	1.241	0.975	0.975	0.973	0.974
oct(ols)	0.955	0.893	0.891	0.893	0.888	0.975	0.937	0.936	0.936	0.935
oct($struc$)	1.085	0.917	0.915	0.916	0.912	1.027	0.943	0.942	0.943	0.942
oct($wlsv$)	1.132	0.933	0.929	0.931	0.927	1.050	0.951	0.949	0.950	0.949
oct($bdshr$)	1.047	0.904	0.897	0.897	0.891	1.009	0.936	0.933	0.934	0.931

Partly bottom up

$\widetilde{\boldsymbol{X}}$ with $\mathrm{ct}(rec_{cs}, bu_{te})$

$$\begin{bmatrix} \widetilde{\boldsymbol{U}}^{[m]}_{te(bu)} & \dots & \widetilde{\boldsymbol{U}}^{[k_2]}_{te(bu)} & \widetilde{\boldsymbol{U}}^{[1]}_{cs(rec)} \\ \widetilde{\boldsymbol{B}}^{[m]}_{te(bu)} & \dots & \widetilde{\boldsymbol{B}}^{[k_2]}_{te(bu)} & \widetilde{\boldsymbol{B}}^{[1]}_{cs(rec)} \end{bmatrix}$$

Temporal

$\widetilde{\boldsymbol{X}}$ with $\mathrm{ct}(rec_{te}, bu_{cs})$

$$\begin{bmatrix} \widetilde{\boldsymbol{U}}^{[m]}_{cs(bu)} & \dots & \widetilde{\boldsymbol{U}}^{[k_2]}_{cs(bu)} & \widetilde{\boldsymbol{U}}^{[1]}_{cs(bu)} \\ \widetilde{\boldsymbol{B}}^{[m]}_{te(rec)} & \dots & \widetilde{\boldsymbol{B}}^{[k_2]}_{te(rec)} & \widetilde{\boldsymbol{B}}^{[1]}_{te(rec)} \end{bmatrix}$$

Cross-sectional

- The blue background indicates generating reconciled forecasts along one dimension, while the pink background indicates the forecasts obtained using bottom-up along the other

L Cross-sectionally reconciled forecasts for $k = 1$ ($\widetilde{\boldsymbol{U}}^{[1]}$ and $\widetilde{\boldsymbol{B}}^{[1]}$) followed by temporal bottom-up

R Temporally reconciled forecasts of the cross-sectional bottom time series ($\widetilde{\boldsymbol{B}}^{[k]}$, $k \in \mathcal{K}$) followed by cross-sectional bottom-up

Cross-temporal covariance approximations

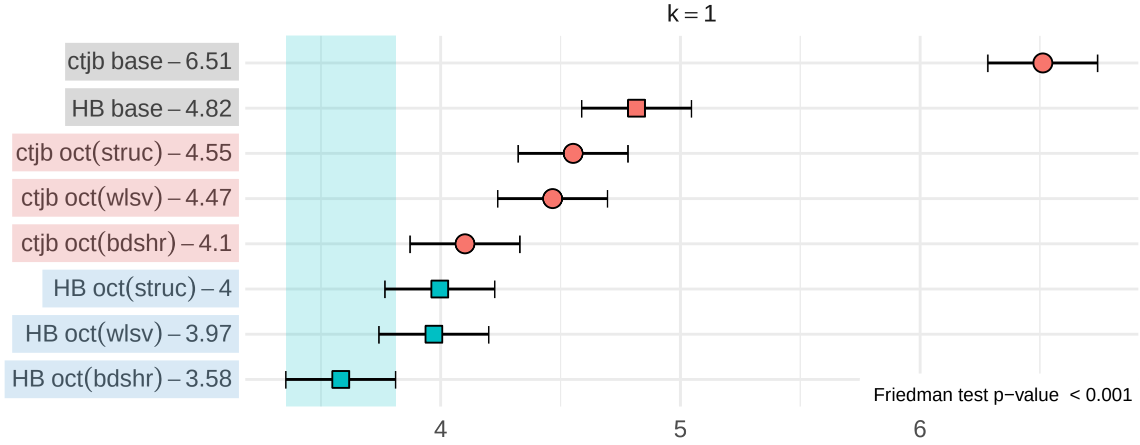
Di Fonzo and Girolimetto (2023a)

- oct(ols) - identity: $\boldsymbol{\Omega}_{ct} = \boldsymbol{I}_{n(k^*+m)}$
- oct($struc$) - structural: $\boldsymbol{\Omega}_{ct} = \mathrm{diag}(\boldsymbol{F}\boldsymbol{1}_{mn_b})$
- oct($wlsv$) - series variance scaling: $\boldsymbol{\Omega}_{ct} = \widehat{\boldsymbol{\Omega}}_{ct,wlsv}$, a straightforward extension of the series variance scaling matrix presented by Athanasopoulos *et al.* (2017)
- oct($bdshr$) - block-diagonal shrunk cross-covariance scaling: $\boldsymbol{\Omega}_{ct} = \boldsymbol{P}\widehat{\boldsymbol{W}}_{ct,shr}^{BD}\boldsymbol{P}'$, with $\widehat{\boldsymbol{W}}_{ct,shr}^{BD}$ a block diagonal matrix where each $k-$block $(k = m, k_{p-1}, \dots, 1)$ is $\boldsymbol{I}_{M_k} \otimes \widehat{\boldsymbol{W}}_{shr}^{[k]}$, $\widehat{\boldsymbol{W}}_{shr}^{[k]}$ is the shrunk estimate of the cross-sectional covariance matrix proposed by Wickramasuriya *et al.* (2019), and $\boldsymbol{P}$ is the commutation matrix such that $\boldsymbol{P}\mathrm{vec}(\boldsymbol{Y}_\tau) = \mathrm{vec}(\boldsymbol{Y}_\tau')$
- oct(shr) - MinT-shr: $\boldsymbol{\Omega}_{ct} = \hat{\lambda}\widehat{\boldsymbol{\Omega}}_{ct,D} + (1-\hat{\lambda})\widehat{\boldsymbol{\Omega}}_{ct}$, where $\hat{\lambda}$ is an estimated shrinkage coefficient (Ledoit and Wolf, 2004), $\widehat{\boldsymbol{\Omega}}_{ct,D} = \boldsymbol{I}_{n(k^*+m)} \odot \widehat{\boldsymbol{\Omega}}_{ct}$ with $\odot$ denoting the Hadamard product, and $\widehat{\boldsymbol{\Omega}}_{ct}$ is the covariance matrix of the cross-temporal one-step ahead in-sample forecast errors

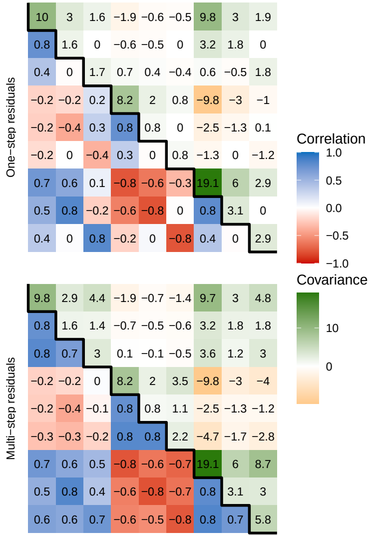
TAFS

Gaussian vs bootstrap MCB Nemenyi test

Comparison between base and the best reconciled forecasts using gaussian □ and bootstrap ○ approaches



From one- to multi-step residuals



- Model residuals may be used to estimate the covariance matrix for the base forecasts $\boldsymbol{\Omega}$ or for the reconciled formula $\boldsymbol{\Omega}_{ct}$
- In time series analysis, it is common to use residuals corresponding to one-step ahead forecasts
- Due to the temporal dimension, residuals corresponding to different forecast horizons are required
- **Multi-step residuals**

$$e_{i,h,j}^{[k]} = x_{i,j+h}^{[k]} - \widehat{x}_{i,j+h|j}^{[k]},$$

where $\widehat{x}_{i,j+h|t}^{[k]}$ is the h-step fitted value.

- In general, these residuals will be **autocorrelated** except when $h = 1$ (one-step residuals)

Dealing with negativity issues: set-negative-to-zero

- One issue in working with time series data is the presence of negative values, which can cause difficulties for certain types of models or analyses
- Cross-sectional non negative reconciliation: Wickramasuriya *et al.* (2020)
- Non negative cross-temporal reconciliation proposed by Di Fonzo and Girolimetto (2022, 2023b)
  - State-of-the-art numerical optimization procedure (osqp, Stellato *et al.*, 2020)
  - Simple heuristic strategy: set-negative-to-zero (sntz)
    1. Negative high-frequency bottom ts reconciled forecasts are set to zero $\rightarrow \widetilde{\boldsymbol{b}}_0^{[1]}$
    2. Apply cross-temporal bottom-up to obtain fully coherent non negative forecasts

$$\underset{\text{Incoherent}}{\widehat{\boldsymbol{x}}} \quad \rightarrow \quad \underset{\text{Coherent}}{\widetilde{\boldsymbol{x}}} \quad \rightarrow \quad \underset{\text{Not negative and coherent}}{\widetilde{\boldsymbol{x}}_0 = \boldsymbol{F}\widetilde{\boldsymbol{b}}_0^{[1]}}$$

+ Sntz requires much less time and computational effort than optimization
+ In the empirical application, both sntz and osqp produce similar quality forecasts

TAFS