Cross-temporal Probabilistic Forecast Reconciliation

Daniele Girolimetto
Department of Statistical Sciences, University of Padova
Email: daniele.girolimetto@phd.unipd.it

George Athanasopoulos
Department of Econometrics and Business Statistics, Monash University
Email: george.athanasopoulos@monash.edu

Tommaso Di Fonzo
Department of Statistical Sciences, University of Padova
Email: tommaso.difonzo@unipd.it

Rob J Hyndman
Department of Econometrics and Business Statistics, Monash University
Email: rob.hyndman@monash.edu

17 March 2023

Abstract

Forecast reconciliation is a post-forecasting process that maps a set of incoherent forecasts into coherent forecasts which satisfy a given set of linear constraints for a multivariate time series. In this paper we extend the state-of-the-art cross-sectional probabilistic forecast reconciliation to the cross-temporal framework, where temporal constraints are also considered. We develop a parametric Gaussian and a non parametric bootstrap approach to draw samples from an incoherent cross-temporal distribution. Multi-step residuals are used for improving the estimation of the covariance matrix, specifically in the time dimension where the usual one-step residuals fail. To address high-dimensionality issues, we propose four alternatives for the covariance matrix by exploiting the two-fold nature (cross-sectional and temporal) of the cross-temporal structure and introduce the idea of overlapping residuals. We investigate the theoretical and empirical proprieties of the proposed approaches via a detailed simulation study. Finally, we consider two empirical forecasting experiments using the Australian GDP and the Australian Tourism Demand datasets to evaluate the feasibility and the performance of the proposed procedures. For both applications, the optimal cross-temporal reconciliation approaches significantly outperform the incoherent base forecasts in terms of the Continuous Ranked Probability Score and the Energy Score. Overall, the paper expands and unifies the notation for cross-sectional, temporal and cross-temporal reconciliation, and investigates the probabilistic cross-temporal framework in more detail. These results demonstrate the effectiveness of the cross-temporal forecast reconciliation approaches in improving the accuracy of probabilistic forecasting models.

Keywords: Multivariate time series, Temporal aggregation, Linear constraints, Coherent, GDP, Tourism flows

1 Introduction

Forecast reconciliation is a post-forecasting process intended to improve the quality of forecasts for a system of linearly constrained multiple time series (Hyndman et al. 2011, Panagiotelis et al. 2021, Di Fonzo & Girolimetto 2022*d*). There are many fields where forecast reconciliation is useful, such as when forecasting GDP and its components, electricity demand and power generation, demand in supply chains with product categories, tourist flows across geographic regions and travel purpose, and more. Moreover, effective decision-making depends on the support of accurate and coherent forecasts.

Classical reconciliation methods addressed the issue of incoherent forecasts in a cross-sectional hierarchy by forecasting only one level and using these to generate forecast on the rest of the structure. The bottom-up approach (Dunn et al. 1976) starts by generating forecasts at the most disaggregate level and summing these to arrive at the desired forecasts for aggregate levels. On the other hand, the top-down approach (Gross & Sohl 1990) forecasts the most aggregated level and then disaggregates it to lower levels (Fliedner 2001, Athanasopoulos et al. 2009). The middle-out method (Athanasopoulos et al. 2009) combines both approaches by selecting an intermediate level and applies top-down for lower levels and bottom-up for upper levels.

These approaches ignore useful information available at other levels (Pennings & van Dalen 2017). Consequently, in the last decade, hierarchical forecasting has significantly evolved to include modern least squares-based reconciliation techniques in the cross-sectional framework (Hyndman et al. 2011, Wickramasuriya et al. 2019, Panagiotelis et al. 2021), later extended to temporal hierarchies (Athanasopoulos et al. 2017, Nystrup et al. 2020). Obtaining coherent forecasts across both the cross-sectional and temporal dimensions (known as *cross-temporal coherence*) has been limited to sequential approaches that address each dimension separately (Kourentzes & Athanasopoulos 2019, Yagli et al. 2019, Punia et al. 2020, Spiliotis et al. 2020). Recently, Di Fonzo & Girolimetto (2023*a*) suggested a unified reconciliation step that takes into account both the cross-sectional and temporal dimensions, instead of dealing with them separately, utilizing the entire cross-temporal hierarchy.

However, these cross-temporal works focus on point forecasting, and do not consider distributional or probabilistic forecasts (Gneiting & Katzfuss 2014). In the cross-sectional framework, there has been some developments towards probabilistic forecasting including Ben Taieb et al. (2017), Panamtash & Zhou (2018), Jeon et al. (2019), Ben Taieb et al. (2021), Corani et al. (2021), Corani et al. (2022), Zambon et al. (2022) and Wickramasuriya (2023). Panagiotelis et al. (2023) made a significant contribution by formalizing cross-sectional probabilistic reconciliation using the geometric framework for point forecast reconciliation proposed by Panagiotelis et al. (2021). They show how a reconciled forecast can be constructed from an arbitrary base forecast when the density of the base forecast is available and when only a sample can be drawn. They also show that in the case of elliptical distributions, the correct predictive distribution can be recovered via linear reconciliation, regardless of the

base forecast location and scale parameters, and derive conditions for this to hold in the special case of reconciliation via projection.

In this paper, we extend cross-sectional probabilistic reconciliation to the cross-temporal case, working on issues related to the two-fold nature of this framework. First, we revise and develop the notation proposed by Di Fonzo & Girolimetto (2023*a*) to generalize the work of Panagiotelis et al. (2023). This allows us to move from cross-temporal point reconciliation to a probabilistic setting through the generalization of definitions and theorems well established in the cross-sectional framework. Second, we propose effective and practical solutions to draw a sample from the base forecast distribution according to either a parametric approach that assumes Gaussianity or a non-parametric approach that bootstraps the base model residuals. Third, we propose some solutions to specific problems that arise when combining the cross-sectional and temporal dimensions. We propose using multi-step residuals to estimate the relationships between different forecast horizons when we deal with temporal levels, since one-step residuals are not suitable for this purpose. to solve high-dimensionality issues we introduce the idea of overlapping residuals and consider alternative forms for constructing the covariance matrix. Fourth, we propose new shrinkage procedures for reconciliation that aim to identify a feasible cross-temporal structure. The methodological contributions described in this paper are implemented in the `FoReco` package for R (Girolimetto & Di Fonzo 2023).

The remainder of the paper is structured as follows. In Section 2, we provide a unified notation for the cross-sectional, temporal and cross-temporal point reconciliation. We generalize the cross-sectional definitions and theorems developed by Panagiotelis et al. (2023) in Section 3, and propose both a parametric Gaussian and a non-parametric bootstrap approach to draw a sample from the base forecast distribution. In Section 4, we analyze the structure of the cross-temporal covariance matrix, proposing four alternative forms, and propose shrinkage approaches for reconciliation. In addition, we explore cross-temporal residuals (overlapping and multi-step) looking at their advantages and limitations. A simulation study is performed in Section 5, to better understand the properties of the methodology. Two empirical applications using the Australian GDP and the Australian Tourism Demand datasets are considered in Sections 6 and 7, respectively. Finally, Section 8 presents conclusions and a future research agenda on this and other related topics.

2 Notation and definitions

Let $\boldsymbol{y}_t = [y_{1,t}, \dots, y_{i,t}, \dots, y_{n,t}]'$ be an n-variate linearly constrained time series observed at the most temporally disaggregated level, with a seasonality of period m (e.g., $m = 12$ for monthly data, $m = 4$ for quarterly data, $m = 24$ for hourly data). Suppose that the constraints are expressed by linear equations such that (Di Fonzo & Girolimetto 2023*a*)

$$\boldsymbol{C}_{cs}\boldsymbol{y}_t = \mathbf{0}_{(n_a \times 1)}, \qquad t = 1, \ \dots, \ T, \tag{1}$$

where $\boldsymbol{C}_{cs}$ is the $(n_a \times n)$ zero constraints cross-sectional matrix, that can be seen as the coefficient matrix of a linear system with n_a equations and n variables.

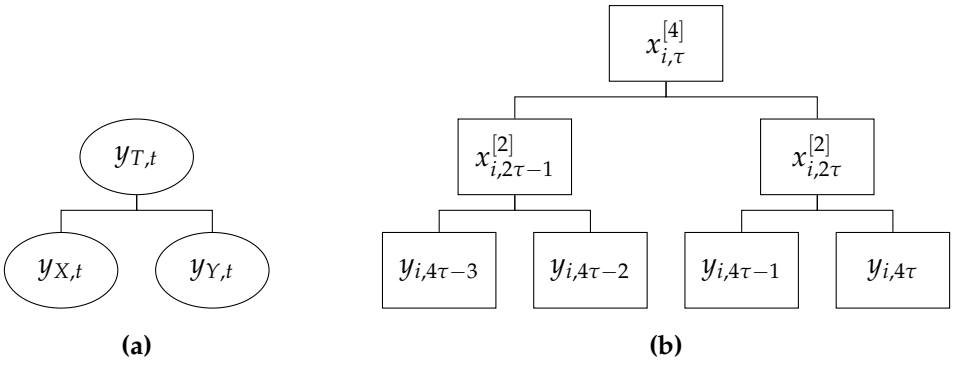

Figure 1: ***(a)*** *A simple two-level cross-sectional hierarchy for 3 linearly constrained time series with $n = 3$, $n_a = 1$ and $n_b = 2$.* ***(b)*** *A temporal hierarchy for a quarterly series ($m = 4$ and $\mathcal{K} = \{4, 2, 1\}$).*

An example is a hierarchical time series where series at upper levels can be expressed by appropriately summing part or all of the series at the bottom level. For example, Figure 1(a) shows the two-level hierarchical structure for three linearly constrained time series such that $y_{T,t} = y_{X,t} + y_{Y,t}$, $\forall t = 1, ..., T$. Now let $\boldsymbol{y}_t = [\boldsymbol{u}_t', \ \boldsymbol{b}_t']'$, where $\boldsymbol{u}_t = [y_{1,t}, \dots, y_{n_a,t}]'$ is the n_a-vector of upper levels time series and $\boldsymbol{b}_t = [y_{(n_a+1),t}, \dots, y_{n,t}]'$ is the n_b-vector of bottom level time series with $n = n_a + n_b$. The upper and lower level time series are connected by the cross-sectional aggregation matrix $\boldsymbol{A}_{cs}$ such that $\boldsymbol{u}_t = \boldsymbol{A}_{cs}\boldsymbol{b}_t$. Following Di Fonzo & Girolimetto (2022*d*), we can always construct a zero-constraints cross-sectional matrix from the aggregation matrix, $\boldsymbol{C}_{cs} = [\boldsymbol{I}_{n_a} \ \ -\boldsymbol{A}_{cs}]$. Finally, the cross-sectional structural matrix is given by $\boldsymbol{S}_{cs} = [\boldsymbol{A}' \ \ \boldsymbol{I}_{n_b}]'$, providing the structural representation (Hyndman et al. 2011)

$$\boldsymbol{y}_t = \boldsymbol{S}_{cs}\boldsymbol{b}_t.$$

Considering the hierarchical example in Figure 1(a), we have

$$\boldsymbol{A}_{cs} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \boldsymbol{C}_{cs} = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{S}_{cs} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In general there is no reason for $\boldsymbol{u}_t$ to be restricted to simple sums of $\boldsymbol{b}_t$; therefore $\boldsymbol{A}_{cs} \in \mathbb{R}^{n_a \times n_b}$ may contain any real values, and not only 0s and 1s.

Considering now the temporal framework, we denote as $\mathcal{K} = \{k_p, k_{p-1}, \dots, k_2, k_1\}$ the set of p factors of m, in descending order, where $k_1 = 1$ and $k_p = m$ (Athanasopoulos et al. 2017). Given a factor k of m, and assuming that $T = Nm$ (where N is the length of the most temporally aggregated version of the series), we can construct a temporally aggregated version of the time series of a single variable $\{\boldsymbol{y}_{i,t}\}_{t=1,\dots,T}$, through the non-overlapping sums of its k successive values, which has a seasonal period equal to $M_k = m/k$:

$$x^{[k]}_{i,j} = \sum_{t=(j-1)k+1}^{jk} y_{i,t}, \qquad j = 1, \dots, T/k, \qquad i = 1, \dots, n.$$

Note that $x^{[1]}_{i,j} = y_{i,t}$. Define τ as the observation index of the most aggregate level k_p. For a fixed temporal aggregation order $k \in \mathcal{K}$, we stack the observations in the column vector

$$\boldsymbol{x}^k_{i,\tau} = \begin{bmatrix} x^{[k]}_{i,M_k(\tau-1)+1} & x^{[k]}_{i,M_k(\tau-1)+2} & \dots & x^{[k]}_{i,M_k\tau} \end{bmatrix}',$$

and obtain the vector for all the temporal aggregation orders

$$\boldsymbol{x}_{i,\tau} = \left[x_{i,\tau}^{[k_p]} \quad \boldsymbol{x}_{i,\tau}^{[k_{p-1}]\prime} \quad \dots \quad \boldsymbol{x}_{i,\tau}^{[1]\prime} \right]', \qquad \tau = 1, \dots, N.$$

The structural representation of the temporal hierarchy (Athanasopoulos et al. 2017) is then

$$\boldsymbol{x}_{i,\tau} = \boldsymbol{S}_{te} \boldsymbol{x}_{i,\tau}^{[1]},$$

where $\boldsymbol{S}_{te} = [\boldsymbol{A}_{te}' \;\; \boldsymbol{I}_m]'$ is the $[(m + k^*) \times m]$ temporal structural matrix,

$$\boldsymbol{A}_{te} = \begin{bmatrix} & \boldsymbol{1}_{k_p}' & \\ \boldsymbol{I}_{\frac{m}{k_{p-1}}} & \otimes & \boldsymbol{1}_{k_{p-1}}' \\ & \vdots & \\ \boldsymbol{I}_{\frac{m}{k_2}} & \otimes & \boldsymbol{1}_{k_2}' \end{bmatrix}$$

is the $(k^* \times m)$ temporal aggregation matrix with $k^* = \sum_{k \in \mathcal{K} \setminus \{k_1\}} \frac{m}{k}$, and $\otimes$ is the Kronecker product. For each series $x_{i,\tau}$, $i = 1, \dots, n$, we have also the zero-constrained representation

$$\boldsymbol{C}_{te} \boldsymbol{x}_{i,\tau} = \boldsymbol{0}_{[k^* \times (m+k^*)]}, \qquad \tau = 1, \dots, N, \qquad i = 1, \dots, n \tag{2}$$

where $\boldsymbol{C}_{te} = [\boldsymbol{I}_{k^*} \;\; -\boldsymbol{A}_{te}]$ is the $[k^* \times (m + k^*)]$ zero constraints temporal matrix. Figure 1(b) shows the hierarchical representation of a quarterly time series, for which $m = 4$, $\mathcal{K} = \{4, 2, 1\}$, and

$$\boldsymbol{A}_{te} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{C}_{te} = \begin{bmatrix} 1 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{S}_{te} = \begin{bmatrix} \boldsymbol{A}_{te} \\ \boldsymbol{I}_4 \end{bmatrix}.$$

When we temporally aggregate each series, the cross-sectional constraints for the most temporally disaggregated series (1) hold for all the temporal aggregation orders

$$\boldsymbol{C}_{cs} \boldsymbol{x}_t^{[k]} = \boldsymbol{0}_{(n_a \times 1)}, \quad \text{for} \quad k \in \mathcal{K} \quad \text{and} \quad t = 1, \dots, T, \tag{3}$$

where $\boldsymbol{x}_t^{[k]} = \left[\boldsymbol{u}_t^{[k]\prime}, \; \boldsymbol{b}_t^{[k]\prime} \right]'$ with $\boldsymbol{u}_t^{[k]} = \left[x_{1,t}^{[k]}, \dots, x_{n_a,t}^{[k]} \right]'$ is the n_a-vector of upper time series and $\boldsymbol{b}_t^{[k]} = \left[x_{(n_a+1),t}^{[k]}, \dots, x_{n,t}^{[k]} \right]'$ is the n_b-vector of bottom time series in the temporal hierarchy.

To include both cross-sectional and temporal constraints at the same time in a unified framework, we stack the series into a $[n \times (m + k^*)]$ matrix $\boldsymbol{X}_\tau$, whose rows and columns represent, respectively, the cross-sectional and the temporal dimension:

$$\boldsymbol{X}_\tau = \begin{bmatrix} \boldsymbol{x}_{1,\tau}' \\ \vdots \\ \boldsymbol{x}_{n,\tau}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_\tau^{[k_p]} & \dots & \boldsymbol{X}_\tau^{[k_1]} \end{bmatrix} \quad \text{with} \quad \boldsymbol{X}_\tau^{[k]} = \begin{bmatrix} \boldsymbol{U}_\tau^{[k]} \\ \boldsymbol{B}_\tau^{[k]}, \end{bmatrix}, \tag{4}$$

where for any fixed k, $\boldsymbol{U}_\tau^{[k]}$ is the $(n_a \times T/k)$ matrix grouping the upper time series, $\boldsymbol{B}_\tau^{[k]}$ is the $(n_b \times T/k)$ matrix grouping the bottom time series. Further,

$$\boldsymbol{C}_{cs} \boldsymbol{X}_\tau = \boldsymbol{0}_{[n_a \times (m+k^*)]} \qquad \text{and} \qquad \boldsymbol{C}_{te} \boldsymbol{X}_\tau' = \boldsymbol{0}_{(k^* \times n)}.$$

We can consider the cross-temporal framework as a generalization of the cross-sectional and temporal frameworks, that simultaneously takes into account both types of constraints. The cross-sectional reconciliation approach proposed by Hyndman et al. (2011) can be obtained by assuming $m = 1$, while the temporal one (Athanasopoulos et al. 2017) is obtained when $n = 1$ (with $n_a = 0$ and $n_b = 1$).

Di Fonzo & Girolimetto (2023*a*) show that the cross-temporal constraints working on the complete set of observations corresponding to time period τ can be expressed in a zero-constrained representation through the full rank $[(n_a m + nk^*) \times n(m + k^*)]$ zero constraints cross-temporal matrix $\boldsymbol{C}_{ct}$ such that

$$\boldsymbol{C}_{ct} = \begin{bmatrix} \boldsymbol{C}_* \\ \boldsymbol{I}_n \otimes \boldsymbol{C}_{te} \end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{C}_{ct}\boldsymbol{x}_\tau = \boldsymbol{0}_{[(n_a m+nk^*)\times 1]} \quad \text{for} \quad \tau = 1, \dots, N, \tag{5}$$

where $\boldsymbol{x}_\tau = \text{vec}(\boldsymbol{X}'_\tau) = [\boldsymbol{x}'_{1,\tau}, \ \dots, \ \boldsymbol{x}'_{n,\tau}]'$, $\boldsymbol{C}_* = [\boldsymbol{0}_{(n_a m \times nk^*)} \ \ \boldsymbol{I}_m \otimes \boldsymbol{C}_{cs}]\boldsymbol{P}'$, and $\boldsymbol{P}$ is the commutation matrix (Magnus & Neudecker 2019, p. 54) such that $\boldsymbol{P}\text{vec}(\boldsymbol{Y}_\tau) = \text{vec}(\boldsymbol{Y}'_\tau)$. A structural representation can be considered as well:

$$\boldsymbol{x}_\tau = \boldsymbol{S}_{ct}\boldsymbol{b}^{[1]}_\tau = s(\boldsymbol{b}^{[1]}_\tau),$$

where

$$\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te} \tag{6}$$

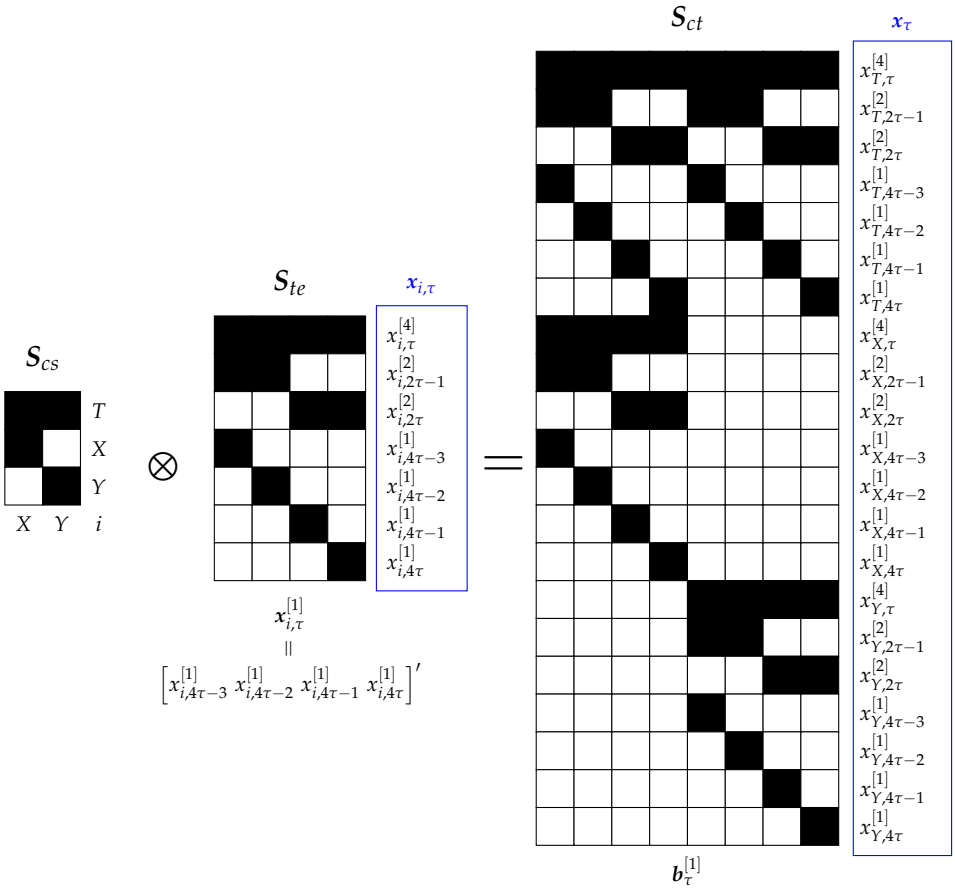

Figure 2: *Representation of the cross-temporal summing matrix $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te}$ defined in equation (6) for a system of 3 linearly constrained quarterly time series (see Figure 1). The white cells indicate a zero, while the black ones a non-zero value.*

is the $[n(k^*+m)\times n_b m]$ cross-temporal summation matrix, $s:\mathbb{R}^{n_b m}\to\mathbb{R}^{n(m+k^*)}$ is the operator describing the pre-multiplication by $\boldsymbol{S}_{ct}$, and $\boldsymbol{b}_\tau^{[1]}=\text{vec}(\boldsymbol{B}_\tau^{[1]\prime})$. We observe that, in agreement with Panagiotelis et al. (2021), $\boldsymbol{x}_\tau$ lies in an $(n_b m)$-dimensional subspace $\mathfrak{s}_{ct}$ of $\mathbb{R}^{n(k^*+m)}$, which we refer to as the *cross-temporal coherent subspace*, spanned by the columns of $\boldsymbol{S}_{ct}$. In Figure 2, we have represented $\boldsymbol{S}_{ct}$ for 3 linearly constrained quarterly time series, shown in Figure 1.

2.1 Optimal point forecast reconciliation

Let $\widehat{\boldsymbol{x}}_h=\text{vec}(\widehat{\boldsymbol{X}}_h')$, $h=1,\dots,H$, be the h-step ahead base forecasts (however obtained) with error covariance matrix given by $\boldsymbol{W}_h=\text{Var}(\widehat{\boldsymbol{x}}_h-\boldsymbol{x})$, where H is the forecast horizon for the most temporally aggregated time series. Denote

$$\widehat{\boldsymbol{X}}_h=\begin{bmatrix}\widehat{\boldsymbol{x}}_{1,h}\\ \vdots\\ \widehat{\boldsymbol{x}}_{n,h}\end{bmatrix}=\begin{bmatrix}\widehat{\boldsymbol{U}}_h^{[m]} & \dots & \widehat{\boldsymbol{U}}_h^{[k]} & \dots & \widehat{\boldsymbol{U}}_h^{[1]}\\ \widehat{\boldsymbol{B}}_h^{[m]} & \dots & \widehat{\boldsymbol{B}}_h^{[k]} & \dots & \widehat{\boldsymbol{B}}_h^{[1]}\end{bmatrix},$$

where $\widehat{\boldsymbol{U}}_h^{[k]}$ is the $(n_a\times M_k)$ matrix grouping the upper time series and $\widehat{\boldsymbol{B}}_h^{[k]}$ is the $(n_b\times M_k)$ matrix grouping the bottom time series for a given temporal aggregation order k. The matrix $\widehat{\boldsymbol{X}}_h$, organized as $\boldsymbol{X}_\tau$ in expression (4), contains incoherent forecasts, such as

$$\boldsymbol{C}_{ct}\widehat{\boldsymbol{x}}_h\neq\boldsymbol{0}_{[(n_a m+nk^*)\times 1]}\qquad h=1,\dots,H,$$

with $\widehat{\boldsymbol{x}}_h=\text{vec}(\widehat{\boldsymbol{X}}_h')$. In this framework, the definition for forecast reconciliation in the cross-sectional framework given by Panagiotelis et al. (2021) can be generalized as follows.

Definition 2.1. Forecast reconciliation aims to adjust the base forecast $\widehat{\boldsymbol{x}}_h$ by finding a mapping $\psi:\mathbb{R}^{n(m+k^*)}\to\mathfrak{s}$ such that $\widetilde{\boldsymbol{x}}_h=\psi(\widehat{\boldsymbol{x}}_h)$, where $\widetilde{\boldsymbol{x}}_h\in\mathfrak{s}$ is the vector of the reconciled forecasts.

For a given forecast horizon $h=1,\dots,H$, the mapping ψ may be defined as a projection onto $\mathfrak{s}$ given by (Panagiotelis et al. 2021, Di Fonzo & Girolimetto 2023*a*)

$$\widetilde{\boldsymbol{x}}_h=\psi(\widehat{\boldsymbol{x}}_h)=\boldsymbol{M}\widehat{\boldsymbol{x}}_h, \tag{7}$$

where $\boldsymbol{M}=\boldsymbol{I}_{n(m+k^*)}-\boldsymbol{\Omega}_{ct}\boldsymbol{C}_{ct}'\left(\boldsymbol{C}_{ct}\boldsymbol{\Omega}_{ct}\boldsymbol{C}_{ct}'\right)^{-1}\boldsymbol{C}_{ct}$, for a positive definite matrix $\boldsymbol{\Omega}_{ct}$, and $\widetilde{\boldsymbol{x}}_h=\text{vec}(\widetilde{\boldsymbol{X}}_h')$. According to Wickramasuriya et al. (2019) showed that the minimum variance linear unbiased reconciled forecasts, satisfying the unbiased condition $\text{E}(\widetilde{\boldsymbol{x}}_h-\boldsymbol{x}_h)=0$, has solution (7) when $\boldsymbol{\Omega}_{ct}=\boldsymbol{W}_h$.

Alternatively, the cross-temporal reconciled forecasts $\widetilde{\boldsymbol{X}}_h$ may be found according to the structural approach proposed by Hyndman et al. (2011) for the cross-sectional framework, yielding $\widetilde{\boldsymbol{x}}_h=\boldsymbol{S}_{ct}\boldsymbol{G}\widehat{\boldsymbol{x}}_h$ for some matrix $\boldsymbol{G}$. Wickramasuriya et al. (2019) showed that this leads to a solution equivalent to the cross-temporally reconciled forecasts in (7), given by

$$\widetilde{\boldsymbol{x}}_h=\psi(\widehat{\boldsymbol{x}}_h)=(s\circ g)(\widehat{\boldsymbol{x}}_h)=\boldsymbol{S}_{ct}\boldsymbol{G}\widehat{\boldsymbol{x}}_h, \tag{8}$$

where $\boldsymbol{G}=(\boldsymbol{S}_{ct}'\boldsymbol{\Omega}_{ct}^{-1}\boldsymbol{S}_{ct})^{-1}\boldsymbol{S}_{ct}'\boldsymbol{\Omega}_{ct}^{-1}$, and $\boldsymbol{M}=\boldsymbol{S}_{ct}\boldsymbol{G}$. In this case, ψ is the composition of two transformations, say $s\circ g$, where $g:\mathbb{R}^{n(m+k^*)}\to\mathbb{R}^{n_b m}$ is a continuous function. In Appendix A we report some cross-sectional, temporal and cross-temporal approximations for the covariance matrix to be used in (7) and (8).

2.2 Cross-temporal bottom-up forecast reconciliation

The classic bottom-up approach (Dunn et al. 1976, Dangerfield & Morris 1992) simply consists in summing-up the base forecasts of the most disaggregated level in the hierarchy to obtain forecasts of the upper-level series. To reduce the computational cost involved in optimal cross-temporal reconciliation, we may be interested in applying a reconciliation along only one dimension (cross-sectional or temporal) and reconstructing the cross-temporal structure using a partly bottom-up approach (Di Fonzo & Girolimetto 2022*a*, 2023*b*, Sanguri et al. 2022).

Figure 3 provides a visual representation of partly bottom-up in a two-step cross-temporal reconciliation approach. On the left (Figure 3a), we first compute the cross-sectionally reconciled forecasts at the highest frequency ($k = 1$), and then apply temporal bottom-up to obtain coherent cross-temporal forecasts. On the right (Figure 3b), we first compute temporally reconciled forecasts for the most disaggregated cross-sectional level, and then apply the cross-sectional bottom-up. We denote these two-step reconciliation approaches, respectively, as $\text{ct}(rec_{te}, bu_{cs})$, and $\text{ct}(rec_{cs}, bu_{te})$, where '$rec_{te}$' and '$rec_{cs}$' denote a forecast reconciliation approach in the temporal and cross-sectional dimensions and, 'bu_{cs}' and 'bu_{te}' denote using bottom-up in the cross-sectional and temporal dimensions, respectively. It is worth noting that the simple cross-temporal bottom-up approach corresponds to $\text{ct}(bu_{cs}, bu_{te}) = \text{ct}(bu_{te}, bu_{cs}) = \text{ct}(bu)$.

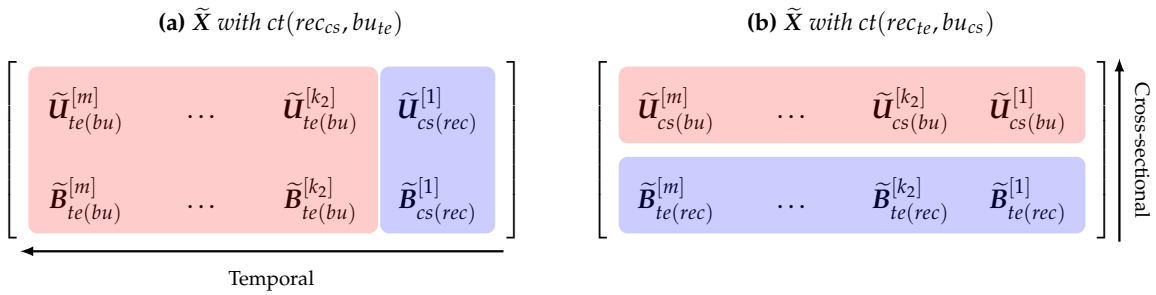

Figure 3: *A visual representation of partly bottom up starting from* (3a) *cross-sectionally reconciled forecasts for the temporal order 1 ($\widetilde{\boldsymbol{U}}^{[1]}$ and $\widetilde{\boldsymbol{B}}^{[1]}$) followed by temporal bottom-up, and* (3b) *temporally reconciled forecasts of the cross-sectional bottom time series ($\widetilde{\boldsymbol{B}}^{[k]}$, $k \in \mathcal{K}$) followed by cross-sectional bottom-up. The blue background indicates generating reconciled forecasts along one dimension, while the pink background indicates the forecasts obtained using bottom-up along the other.*

3 Probabilistic forecast reconciliation

To introduce the idea of coherence and probabilistic forecast reconciliation, we adapt the notations and the formal definitions introduced in Wickramasuriya (2023) and Panagiotelis et al. (2023) for the cross-sectional probabilistic case. These definitions can also be generalized to the cross-temporal framework by following the approach developed by Corani et al. (2022) for count data. However, in this paper we only focus on the continuous case.

Our aim is to extend these definitions to *cross-temporal coherent probabilistic forecasts* and *cross-temporal probabilistic forecast reconciliation*. Let $(\mathbb{R}^{n_b m}, \mathcal{F}_{\mathbb{R}^{n_b m}}, \nu)$ be a probability space for

the bottom time series $\boldsymbol{b}_{\tau}^{[1]}$, where $\mathcal{F}_{\mathbb{R}^{n_b m}}$ is the Borel σ-algebra on $\mathbb{R}^{n_b m}$. Then a σ-algebra $\mathcal{F}_{\mathfrak{s}}$ can be constructed from the collection of sets $s(\mathcal{B})$ for all $\mathcal{B} \in \mathcal{F}_{\mathbb{R}^{n_b m}}$.

Definition 3.1 (Cross-temporal coherent probabilistic forecasts)**.** Given the probability space $(\mathbb{R}^{n_b m}, \mathcal{F}_{\mathbb{R}^{n_b m}}, \nu)$, we define the coherent probability space as the triple $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \check{\nu})$ satisfying the following property:

$$\check{\nu}(s(\mathcal{B})) = \nu(\mathcal{B}), \quad \forall \mathcal{B} \in \mathcal{F}_{\mathbb{R}^{n_b m}}.$$

Let $(\mathbb{R}^{n(m+k^*)}, \mathcal{F}_{\mathbb{R}^{n(m+k^*)}}, \hat{\nu})$ be a probability space referring to the incoherent probabilistic forecast $(\widehat{\boldsymbol{x}}_h)$ for all the n series in the system at any temporal aggregation order $k \in \mathcal{K}$.

Definition 3.2 (Cross-temporal probabilistic forecast reconciliation)**.** The reconciled probability measure of $\hat{\nu}$ with respect to ψ is a probability measure $\tilde{\nu}$ on $\mathfrak{s}$ with σ-algebra $\mathcal{F}_{\mathfrak{s}}$ satisfying

$$\tilde{\nu}(\mathcal{A}) = \hat{\nu}(\psi^{-1}(\mathcal{A})), \quad \forall \mathcal{A} \in \mathcal{F}_{\mathfrak{s}}, \tag{9}$$

where $\psi^{-1}(\mathcal{A}) = \{x \in \mathbb{R}^{n(m+k^*)} : \psi(x) \in \mathcal{A}\}$ denotes the pre-image of $\mathcal{A}$.

The map ψ may be obtained as the composition $s \circ g$, as for the cross-temporal point reconciliation (8). With the following result, we can separate the mechanism used to generate the base forecasts samples from the reconciliation phase.

Theorem 3.1 (Cross-temporal reconciled samples)**.** Suppose that $(\widehat{\boldsymbol{x}}_1, \dots, \widehat{\boldsymbol{x}}_L)$ is a sample drawn from a (cross-temporal) incoherent probability measure $\hat{\nu}$. Then $(\widetilde{\boldsymbol{x}}_1, \dots, \widetilde{\boldsymbol{x}}_L)$, where $\widetilde{\boldsymbol{x}}_\ell = \psi(\widehat{\boldsymbol{x}}_\ell)$ and $\ell = 1, \dots, L$, is a sample drawn from the (cross-temporal) reconciled probability measure $\tilde{\nu}$ defined in (9).

Proof. See Theorem 4.5 from Panagiotelis et al. (2023) using Definition 3.2. □

Theorem 3.1 is the cross-temporal extension of Theorem 4.5 in Panagiotelis et al. (2023). It means that a sample from the reconciled distribution can be obtained by reconciling each member of a sample from the incoherent distribution.

3.1 Parametric framework: Gaussian reconciliation

It is possible to obtain a reconciled probabilistic forecast analytically for some parametric distributions, such as the multivariate normal (Corani et al. 2021, Eckert et al. 2021, Panagiotelis et al. 2023, Wickramasuriya 2023). In the cross-sectional framework, Panagiotelis et al. (2023) show that, starting from an elliptical distribution for the base forecasts, the reconciled forecast distribution is also elliptical. Using the results shown in Section 2, we may extend[1] this results to the cross-temporal case. To obtain a reconciled forecast using the multivariate normal distribution, we start with a base forecast distributed as $\mathcal{N}(\widehat{\boldsymbol{x}}, \boldsymbol{\Sigma})$, where $\widehat{\boldsymbol{x}}_h$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix of the base forecasts. The reconciled forecast distribution is then given by $\mathcal{N}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{\Omega}})$, where

$$\widetilde{\boldsymbol{x}} = \boldsymbol{M}\widehat{\boldsymbol{x}} \quad \text{and} \quad \widetilde{\boldsymbol{\Omega}} = \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}', \tag{10}$$

[1]We assume $H = 1$ and simplify the notation by removing the h suffix without loss of generality

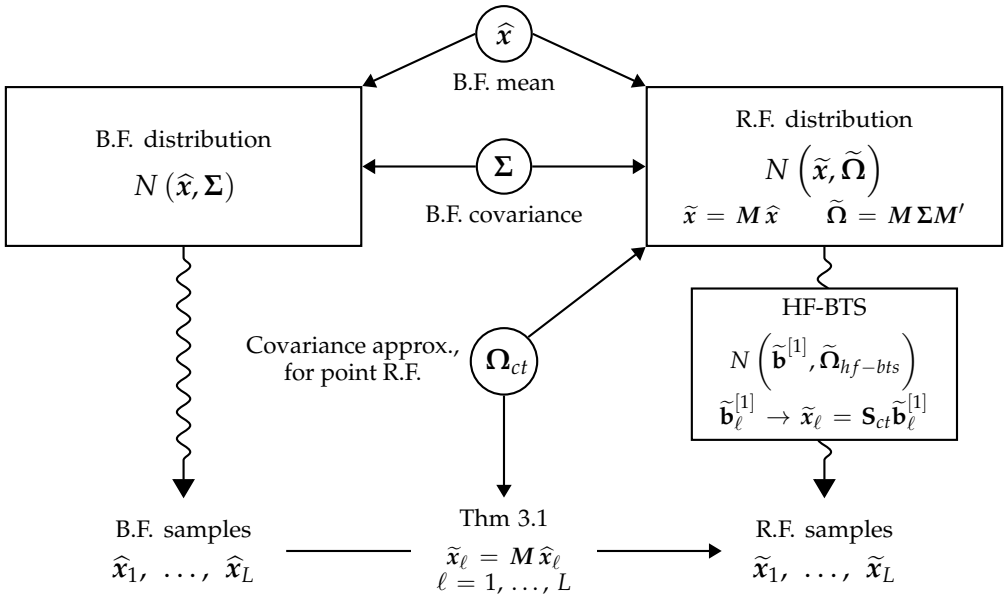

Figure 4: *Visual description of cross-temporal forecast reconciliation in the Gaussian framework, as described in Section 3.1. The acronyms R.F and B.F. stand for Reconciled Forecasts and Base Forecasts, respectively. HF-BTS stands for High Frequency Bottom Time Series.*

where $\boldsymbol{M}$ is the projection matrix defined in (7). Note that if we assume that $\boldsymbol{\Sigma} = \boldsymbol{\Omega}_{ct}$, then the covariance matrix in (10) simplifies to $\widetilde{\boldsymbol{\Omega}} = \boldsymbol{M}\boldsymbol{\Omega}_{ct}$. In the cross-temporal case, sensibly estimating the covariance matrix $\boldsymbol{\Sigma}$ can be difficult because we need to simultaneously consider both the temporal and cross-sectional structures. This requires many parameters to be estimated, which can be challenging in practice. Additionally, naively using one-step residuals to estimate the cross-temporal correlation structure can lead to an inappropriate estimate of the covariance matrix[2]. These challenges will be explored in more depth in the following sections.

Focusing on the computational aspect[3], we can take several steps to reduce the time required to obtain simulations from the reconciled forecast distribution. For example when dealing with a genuine hierarchical structure (that share the same top- and bottom-level variables, Di Fonzo & Girolimetto 2022*d*), it is not necessary to simulate from a normal distribution with a defined covariance matrix for the entire structure. Instead, we can utilize the properties of elliptical distributions to simulate from the high frequency bottom time series and then obtain the complete simulation through the $\boldsymbol{S}_{ct}$ matrix. Furthermore, we do not need to calculate the reconciled mean and variance and generate a new sample if we already have a sample from the normal distribution of the base forecasts; we can simply apply the point forecast reconciliation (7) as outlined in Theorem 3.1. The relationships between base and reconciled forecast distributions and their respective simulations through Theorem 3.1 are depicted in Figure 4.

[2]In particular, some temporal covariances are fixed to zero (see the Figure A.1 in the online appendix for more details).

[3]We use two R packages to sample from a the base forecast gaussian distribution: `Rfast` (Papadakis et al. 2022) when the dimension are very big (Section 7) and `MASS` (Venables & Ripley 2002) otherwise (Sections 5 and 6).

3.2 Non-parametric framework: bootstrap reconciliation

Analytical expressions for the base and reconciled forecast distributions are sometimes challenging to express, and sometimes unrealistic parametric assumptions are used. We propose a procedure called *cross-temporal joint (block) bootstrap* (**ctjb**) to generate samples from the base forecast distributions that preserve cross-temporal relationships. This approach involves drawing samples of all series at the same time from the most temporally aggregated level, and using the most temporally aggregated level to determine the corresponding time indices for the other levels.

Let $\widehat{\boldsymbol{E}}^{[k]}$ be the $(n \times T/k)$ matrix of the residuals for $k \in \mathcal{K}$. Figure 5 (on the left) provides a visualization of these matrices and how they are related to each other for the example in Figure 1. It is assumed that the residuals cover four years ($N = 4$): the green color corresponds to the first year, the blue to the second year, and so on. Further, let $\mathcal{M}_i$ be the model used to calculate the base forecasts and residuals for the i^{th} series. In this work, we assume $\mathcal{M}_i$ to be a univariate model, however nothing prevents the use of multivariate models, perhaps for different temporal levels or for groups of time series.

Assuming $H = 1$, τ is a random draw with replacement from $1, \dots, N$ and the ℓ^{th} bootstrap incoherent sample is

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

where $f_i(\cdot)$ depends on the fitted univariate model $\mathcal{M}_i$. That is, $\widehat{\boldsymbol{x}}^{[k]}_{i,l}$ is a sample path simulated for the i^{th} series with error approximated by the corresponding block bootstrapped sample residual $\widehat{\boldsymbol{e}}^{[k]}_i$, the i^{th} row of

$$\widehat{\boldsymbol{E}}^{[k]}_\tau = \begin{bmatrix} \widehat{e}^{[k]}_{1,M_k(\tau-1)+1} & \cdots & \widehat{e}^{[k]}_{1,M_k\tau} \\ \vdots & \ddots & \vdots \\ \widehat{e}^{[k]}_{n,M_k(\tau-1)+1} & \cdots & \widehat{e}^{[k]}_{n,M_k\tau} \end{bmatrix}.$$

Figure 5 (on the right) shows $\widehat{\boldsymbol{E}}^{[k]}_\tau$, $k \in \{4, 2, 1\}$, for the quarterly cross-temporal hierarchy in Figure 1.

One of the main advantages of the cross-temporal joint bootstrap is that it allows us to accurately account for the dependence between the different levels of temporal aggregation and not only the cross-sectional dependencies. By sampling residuals from the most temporally aggregated level and using it to determine the indices for the other levels, we can ensure that the bootstrap sample reflects the underlying data distribution. Additionally, the cross-temporal joint bootstrap is easy to implement in R (R Core Team 2022) using the package `forecast` (Hyndman et al. 2023) for many forecasting models, making it a practical and efficient tool. Furthermore, this approach is easily scalable in order to utilize multiple computing power simultaneously for each individual series. This can be especially useful when dealing with large datasets or when trying to speed up the analysis process.

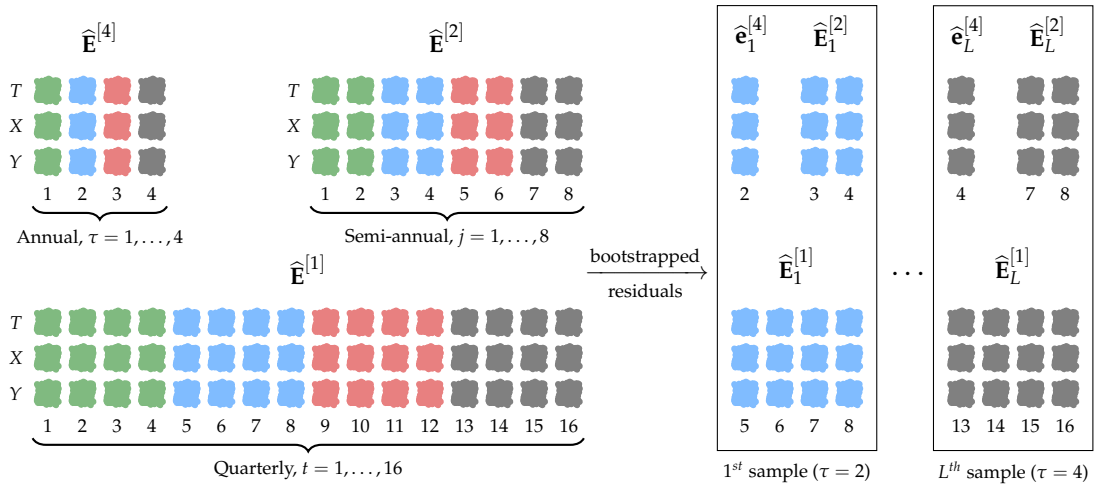

Figure 5: *Example of bootstrapped residuals for 3 linearly constrained quarterly time series (see Figure 1). On the left there are the residual matrices with 4 years of data ($N = 4$): the green color corresponds to the first year, the blue to the second year, the red to the third year and the black to the fourth year. On the right the bootstrapped residuals are represented.*

4 Cross-temporal covariance matrix estimation

As the covariance matrix $\mathbf{\Omega}$ is unknown in practice, a natural estimate is the empirical sample covariance matrix of the base forecasts $\widehat{\mathbf{\Omega}}$. In this section, our focus will be exclusively on the cross-temporal framework., this means that we have to estimate $r = n(k^* + m)[n(k^* + m) - 1]/2$ different parameters. A possible solution to estimating many parameters when we have fewer observations than r, is to construct a shrinkage estimator (Efron 1975, Efron & Morris 1975, 1977), using a convex combination of $\widehat{\mathbf{\Omega}}$ and a diagonal target matrix $\widehat{\mathbf{\Omega}}_D = \widehat{\mathbf{\Omega}} \odot \boldsymbol{I}_{n(k^*+m)}$, such that

$$\widehat{\mathbf{\Omega}}_G = \lambda \widehat{\mathbf{\Omega}}_D + (1 - \lambda)\widehat{\mathbf{\Omega}} \tag{11}$$

where $\lambda \in [0, 1]$ is the shrinkage intensity parameter that can be estimate using the unbiased estimator proposed by Ledoit & Wolf (2004) (see Schäfer & Strimmer 2005). The linear combination involving these two matrices is referred to as *Global shrinkage* (G), where all off-diagonal elements are shrunk towards zero. $\widehat{\mathbf{\Omega}}_G$ corresponds to the matrix used by the reconciliation approach oct(shr) shown in Appendix A.

However, 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. Therefore, we propose to estimate a smaller matrix, and to use the cross-sectional and/or temporal structure to obtain a better estimator for the covariance matrix of the entire system. Given that $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te}$, it is possible to express the actual covariance matrix in terms of three smaller matrices such that (see Appendix B)

$$\begin{aligned} \mathbf{\Omega} &= \boldsymbol{S}_{ct} \mathbf{\Omega}_{hf\text{-}bts} \boldsymbol{S}_{ct}' \\ &= (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})\, \mathbf{\Omega}_{hf}\, (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})' \\ &= (\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})\, \mathbf{\Omega}_{bts}\, (\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})', \end{aligned} \tag{12}$$

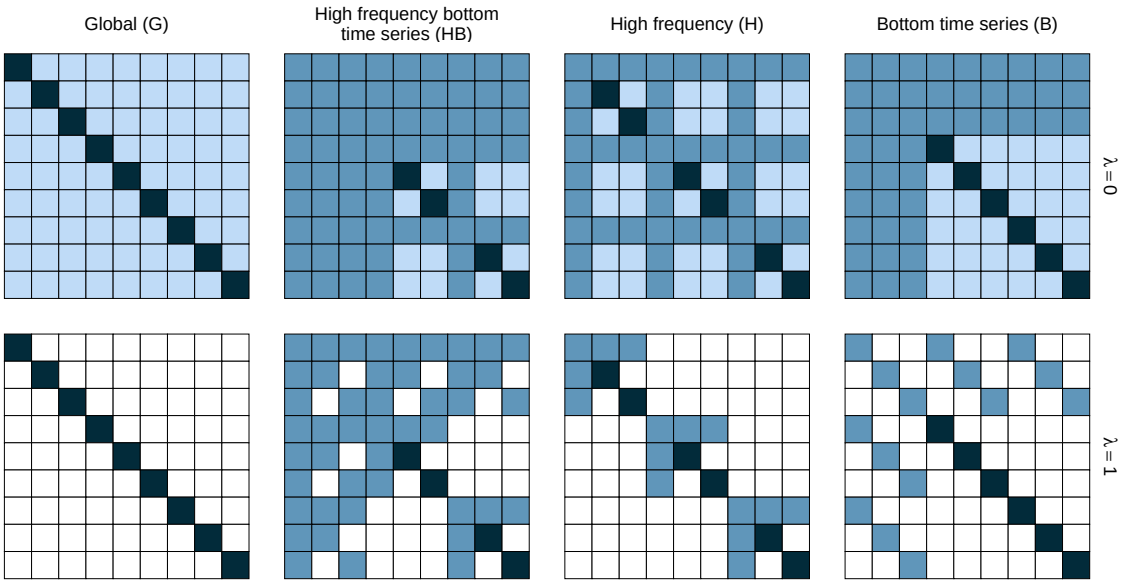

Figure 6: *Representation of four types of covariance matrices that can be obtained from the cross-temporal hierarchical structure (3 time series and $m = 2$) for two different values of $\lambda \in \{0, 1\}$, the shrinkage parameter. The cells that are not modified by shrinkage are colored black, those actively involved in the shrinkage phase are colored light blue, and those derived from and not estimated by the base forecasts errors are colored blue. Additionally, for $\lambda = 1$, the cells corresponding to a zero value are colored white.*

where $\boldsymbol{\Omega}_{\textit{hf-bts}}$ is the $(n_b m \times n_b m]$ covariance matrix for the bottom time series at temporal aggregation level $k = 1$ (highest frequency bottom time series), $\boldsymbol{\Omega}_{\textit{hf}}$ is the $(nm \times nm)$ covariance matrix related to all the high frequency time series and $\boldsymbol{\Omega}_{\textit{bts}}$ is the $[n_b(k^* + m) \times n_b(k^* + m)]$ covariance matrix related to bottom time series at any temporal aggregation.

Therefore, we can apply the idea of "Stein-type shrinkage" (Efron & Morris 1977) to $\boldsymbol{\Omega}_{\textit{hf-bts}}$, $\boldsymbol{\Omega}_{\textit{hf}}$ and $\boldsymbol{\Omega}_{\textit{bts}}$ by using the corresponding empirical base forecasts residuals estimation. We obtain (see Appendix B) the *High frequency Bottom time series shrinkage matrix* (HB)

$$\widehat{\boldsymbol{\Omega}}_{HB} = \lambda \boldsymbol{S}_{ct} \widehat{\boldsymbol{\Omega}}_{\textit{hf-bts},D} \boldsymbol{S}_{ct}' + (1 - \lambda) \boldsymbol{S}_{ct} \widehat{\boldsymbol{\Omega}}_{\textit{hf-bts}} \boldsymbol{S}_{ct}',$$

the *High frequency shrinkage matrix* (H)

$$\widehat{\boldsymbol{\Omega}}_{H} = \lambda (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te}) \widehat{\boldsymbol{\Omega}}_{hf,D} (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})' + (1 - \lambda)(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te}) \widehat{\boldsymbol{\Omega}}_{\textit{hf}} (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})'$$

and the *Bottom time series shrinkage matrix* (B)

$$\widehat{\boldsymbol{\Omega}}_{B} = \lambda \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right) \widehat{\boldsymbol{\Omega}}_{bts,D} \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right)' + (1 - \lambda) \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right) \widehat{\boldsymbol{\Omega}}_{bts} \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right)',$$

where $\widehat{\boldsymbol{\Omega}}_{j,D} = \boldsymbol{I}_{n_b m} \odot \widehat{\boldsymbol{\Omega}}_{j}$ is a diagonal matrix ($j = \{\textit{hf-bts},\ \textit{hf},\ \textit{bts}\}$) and λ is the shrinkage parameter.

Figure 6 gives a visual insights on the covariance matrices obtainable with $\lambda = 0$ and $\lambda = 1$, respectively, for a simple cross-temporal hierarchical structure with 3 time series and $\mathcal{K} = \{2, 1\}$ (e.g, cross-temporal semi-annual, see the Monte Carlo simulation in Section 5). Another important aspect is the number of parameters to be estimated through the residuals

Method	# of different parameters	AR(2)	GDP	Tourism
G	$r = \frac{n(k^*+m)[n(k^*+m)-1]}{2}$	36	221 445	108 052 350
HB	$r_{HB} = \frac{n_b m[n_b m-1]}{2} < r$	6 (83%)	30 876 (86%)	6 655 776 (94%)
H	$r_{HB} < \frac{nm[nm-1]}{2} < r$	15 (58%)	72 390 (67%)	19 848 150 (82%)
B	$r_{HB} < \frac{n_b(k^*+m)[n_b(k^*+m)-1]}{2} < r$	15 (58%)	94 395 (57%)	36 231 328 (66%)

Table 1: *Number of different parameters that need to be estimated for the Monte Carlo simulation (AR(2), see Section 5), the Australian GDP (see Section 6) and the Australian Tourism Demand (see Section 7) forecasting experiments: the first one has* 3 *time series (one upper and two bottom) with temporal aggregation* $\mathcal{K} = \{2, 1\}$*; the second one has* 95 *quarterly (*$m = 4$ *and* $k^* = 3$*) time series (*62 *free and* 33 *constraints, see Di Fonzo & Girolimetto* 2022d*); the last one has a total of 525 monthly (*$m = 12$ *and* $k^* = 16$*) time series (*304 *bottom and* 221 *upper). The percentage reductions in the number of parameters compared to the global approach are reported in parentheses.*

of the base forecasts. In Table 1 we report the number of different parameters for the Monte Carlo simulation (AR2) and for the two forecasting experiment: Australian GDP (see Section 6) and Australian Tourism Demand (see Section 7). In addition, we calculate also the percentage reductions in the number of parameters compared to the global approach, that is: % reduction $= 100(1 - r_i/r_G)$ with $i \in \{HB, H, B\}$. As we can see, G involves a considerable number of parameters compared to other procedures. We observe that HB leads to a decrease of around 85%, whereas the H and B approaches look as a compromise between the previous two. In general, as m and n increase (see Appendix B), using H requires the estimation of a smaller number of parameters.

In our simulations and forecasting experiments, we will be closely analyzing these different constructions with a dual purpose. First, we will use the full covariance matrix ($\lambda = 0$) of the base forecasts to obtain a base forecasts sample of the linearly constrained time series under Gaussianity assumption. Then, we will use the shrinkage versions as approximations of the covariance matrix to be used for reconciliation. This will allow us to better understand the properties and abilities of each parametrization.

4.1 Multi-step residuals

Model residuals may be used to estimate the covariance matrix in cross-temporal forecast reconciliation. In time series analysis, it is common to use residuals corresponding to one-step ahead forecasts, but because of the different temporal dimensions, we need residuals corresponding to different forecast horizons. Thus, we define *multi-step residuals* as

$$e^{[k]}_{i,h,j} = x^{[k]}_{i,j+h} - \widehat{x}^{[k]}_{i,j+h|t}, \quad i = 1, \dots, n \; j = 1, \dots, N_k$$

where $N_k = Nm/k$ and $\widehat{x}^{[k]}_{i,j+h|t}$ is the h-step fitted value, calculated as the h-step-ahead forecast using data up to time j. In general, these residuals will be autocorrelated except when $h = 1$.

Following Di Fonzo & Girolimetto (2023*a*), we use a matrix organization of the residuals similar to the one for the base forecasts in Section 2.1. Specifically, let N be the total number of observations for the most temporally aggregate time series. Then, the N_k-vectors of multi-step residuals for the temporal aggregation k and the series i,

$$\boldsymbol{e}^{[k]}_{i,h} = \begin{bmatrix} e^{[k]}_{i,h,1} & e^{[k]}_{i,h,2} & \dots & e^{[k]}_{i,h,N_k} \end{bmatrix}' \quad h = 1, \dots, \frac{m}{k},$$

can be organized in matrix form as

$$\boldsymbol{E}^{[k]}_{i} = \begin{bmatrix} e^{[k]}_{i,1,1} & e^{[k]}_{i,2,2} & \dots & e^{[k]}_{i,\frac{m}{k},\frac{m}{k}} \\ \vdots & \vdots & & \vdots \\ e^{[k]}_{i,1,N_k-\frac{m}{k}+1} & e^{[k]}_{i,2,N_k-\frac{m}{k}+2} & \dots & e^{[k]}_{i,\frac{m}{k},N_k} \end{bmatrix}. \tag{13}$$

Let $\boldsymbol{E}_i = \begin{bmatrix} \boldsymbol{E}^{[m]}_i & \boldsymbol{E}^{[k_p-1]}_i & \dots & \boldsymbol{E}^{[1]}_i \end{bmatrix}$. Then the $[N \times n(m + k^*)]$ cross-temporal residual matrix is given by

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{E}_1 & \boldsymbol{E}_2 & \dots & \boldsymbol{E}_n \end{bmatrix}. \tag{14}$$

4.2 Overlapping residuals

Another issue that arises in the case of cross-temporal reconciliation is the low number of available residuals, especially for the higher orders of temporal aggregation. A possible solution is to use residuals calculated using overlapping series by allowing the year to have a varying starting time. To better explain how to calculate overlapping residuals, assume we have a single series $\boldsymbol{y} = [y_1 \; y_2 \; y_3 \; \dots \; y_{T-1} \; y_T]'$. We can construct k non overlapping series such that

$$\boldsymbol{x}^{[k],s} = \left\{ x^{[k],s}_j \right\}_{j=1}^{\frac{T}{k}-s} \qquad \text{where} \quad x^{[k],s}_j = \sum_{t=(j-1)k+s+1}^{jk-s} y_t,$$

with $s = 0, \dots, (k-1)$. For example, suppose we have a biannual series with $k = 2$ and $T = 6$, then we can construct two annual time series depending on which time is deemed the start of the year:

$$\boldsymbol{x}^{[2],0} = \begin{bmatrix} x^{[2],0}_1, \; x^{[2],0}_2, \; x^{[2],0}_3 \end{bmatrix}' = \begin{bmatrix} y_1 + y_2, \; y_3 + y_4, \; y_5 + y_6 \end{bmatrix}'$$

and

$$\boldsymbol{x}^{[2],1} = \begin{bmatrix} x^{[2],1}_1, \; x^{[2],1}_2 \end{bmatrix}' = \begin{bmatrix} y_2 + y_3, \; y_4 + y_5 \end{bmatrix}'.$$

To calculate overlapping residuals, we propose the following steps:

- **step 1)** Fit a model to $\boldsymbol{x}^{[k],0}$; i.e., select an appropriate model and estimate the model parameters using the available data.
- **step 2)** Calculate the residuals for $\boldsymbol{x}^{[k],0}$.
- **step 3)** Apply the same model in step 1 to $\boldsymbol{x}^{[k],s}$ for $s = 1, \dots, k-1$, without re-estimating the parameters.

step 4) Calculate the residuals for $\boldsymbol{x}^{[k],s}$ for $s = 1, \ldots, k-1$ using the same method as step 2.

step 5) Organize the residuals from steps 2 and 4 according to the structure in (14).

The resulting residuals can be used to estimate the covariance matrix in cross-temporal forecast reconciliation. Operating in this way increases the number of available residuals, particularly when working with higher frequency observations such as monthly or daily data.

It is important to note that this approach assumes that the model used in step 1 is appropriate for all the different series $\boldsymbol{x}^{[k],s}$. Some seasonal models will not be appropriate as the seasonal pattern will be shifted for different values of s. However, this will not affect seasonal ARIMA models as the seasonality is defined in terms of lags which are unaffected by the value of s.

5 Monte Carlo simulation

We study the effect of combining cross-sectional and temporal aggregations, using a simple hierarchy that allows us to effectively visualize the quantities involved, such as the covariance matrix. Additionally, the small size and nature of the data generating process make it possible to accurately calculate the true cross-temporal covariance structure, which can provide valuable insights into nature the time series data involve in the forecast reconciliation process.

Consider a 2-level hierarchical structure with three time series (one upper series, A, and two bottom series, B and C) such that the cross-sectional aggregation matrix is $\boldsymbol{A}_{cs} = [1 \quad 1]$ ($A = B + C$). The temporal structure we are considering is equivalent to using semi-annual data with $\mathcal{K} = \{2, 1\}$ and $m = 2$. The assumed Data-Generating Processes (DPG) for the semi-annual bottom level series are two AR(2) given by

$$\begin{cases} y_{B,t} = \boldsymbol{\phi}_B' \begin{bmatrix} y_{B,t-1} \\ y_{B,t-2} \end{bmatrix} + \varepsilon_{B,t} \\ y_{C,t} = \boldsymbol{\phi}_C' \begin{bmatrix} y_{C,t-1} \\ y_{C,t-2} \end{bmatrix} + \varepsilon_{C,t} \end{cases}$$

with parameters[4] $\boldsymbol{\phi}_B = [1.34 \ -0.74]'$ and $\boldsymbol{\phi}_C = [0.95 \ -0.42]'$. The error $\boldsymbol{\varepsilon}_t = [\varepsilon_{B,t} \quad \varepsilon_{C,t}]'$ driving the process is drawn from a multivariate normal distribution with standard deviations simulated from a uniform distribution between 0.5 and 2 and a fixed correlation of -0.8. The cross-sectional error covariance matrix is thus given by

$$\boldsymbol{\Omega}_{cs} = \begin{bmatrix} 0.9 & 0 \\ 0 & 1.8 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 1.8 \end{bmatrix} = \begin{bmatrix} \sigma_B^2 & \sigma_{BC} \\ \sigma_{BC} & \sigma_C^2 \end{bmatrix}.$$

To obtain the remaining series, the bottom series are then cross-temporally aggregated.

[4]The $\boldsymbol{\phi}_B$ and $\boldsymbol{\phi}_C$ parameters are estimated from the "Lynx" and "Hare" time series contained in the `pelt` dataset of the `tsibbledata` package for R (O'Hara-Wild et al. 2022).

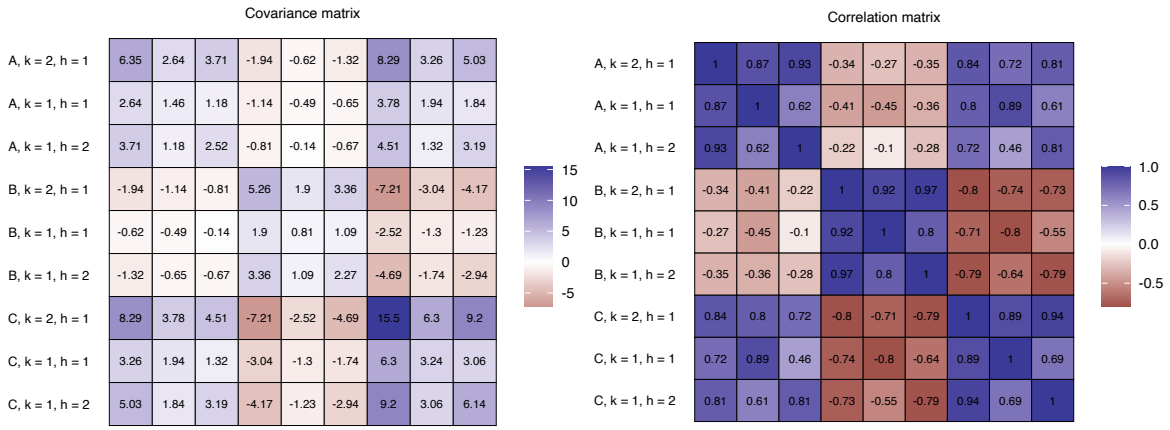

Figure 7: *True cross-temporal covariance (left) and correlation (right) error matrix of the reconciled forecasts with* $\sigma_B = 0.9$, $\sigma_C = 1.8$, $\boldsymbol{\phi}_B = [1.34 \ \ -0.74]'$, $\boldsymbol{\phi}_C = [0.95 \ \ -0.42]'$ *and* $\rho = -0.8$.

For the forecast experiment, the base forecasts are computing using AR models where the order is automatically determined by the algorithm proposed by Hyndman & Khandakar (2008) and implemented in the R package `forecast` (Hyndman et al. 2023), thus allowing for possible mis-specification in the models. The training window length is 500 years, consisting of 1000 high frequency observations. The experiment is replicated 500 times, with a forecast horizon of 1 year.

Since the AR(2) models used as DPG for the bottom series B and C at the most disaggregated temporal level are known, we may compute the true covariance matrix for one-step ahead forecasts at the annual level $\boldsymbol{\Omega}_{ct} = \boldsymbol{S}_{ct}\boldsymbol{\Omega}_{hf\text{-}bts}\boldsymbol{S}'_{ct}$, where

$$\boldsymbol{\Omega}_{hf\text{-}bts} = \begin{bmatrix} \sigma_B^2 & & & \\ \phi_{B,1}\sigma_B^2 & \sigma_B^2\left(1+\phi_{B,1}^2\right) & & \\ \sigma_{BC} & \phi_{B,1}\sigma_{BC} & \sigma_C^2 & \\ \phi_{C,1}\sigma_{BC} & \sigma_{BC}\left(1+\phi_{B,1}\phi_{C,1}\right) & \phi_{C,1}\sigma_C^2 & \sigma_C^2\left(1+\phi_{C,1}^2\right) \end{bmatrix}.$$

The detailed calculations can be found in Appendix C. Figure 7 shows both the covariance matrix and the correlation matrix for fixed parameters.

To construct cross-temporal samples of the base forecasts, we use the Gaussian and bootstrap approaches discussed in Sections 3.1 and 3.2, respectively. For the parametric approach we use multi-step residuals with the different covariance matrix structures analyzed in Section 4, while for the non-parametric approach, we use regular one-step residuals. We do not use overlapping residuals in our analysis as we have the advantage of generating a large number of observation. Ten different reconciliation approaches have been adopted (see Table 2): $\text{ct}(bu)$, $\text{ct}(shr_{cs}, bu_{te})$, $\text{ct}(wlsv_{te}, bu_{cs})$, $\text{oct}(wlsv)$, $\text{oct}(bdshr)$, $\text{oct}_h(shr)$, $\text{oct}_h(bshr)$, $\text{oct}_h(hshr)$ and $\text{oct}_h(hbshr)$.

Label	Description
$\mathrm{ct}(bu)$	Simple cross-temporal bottom-up (Section 2.2).
$\mathrm{ct}(\cdot, bu_{te})$	Partly bottom-up (Section 2.2) starting from cross-sectional reconciled forecasts using the *shr* and *wls* approaches (Table A.1).
$\mathrm{ct}(wlsv_{te}, bu_{cs})$	Partly bottom-up (Section 2.2) starting from temporally reconciled forecasts using the *wlsv* approach (Table A.1).
$\mathrm{oct}(\cdot)$	Optimal cross-temporal reconciliation for the *ols*, *struc*, *wlsv* and *bdshr* approaches (see Appendix A). One-step residuals were used with *wlsv* and *bdshr*.
$\mathrm{oct}_{h}(\cdot)$	Optimal cross-temporal reconciliation with multi-step residuals (see Section 4.1) for the shrinkage approaches presented in Section 4: *shr* stands for *Global shrinkage*, *hshr* for *High frequency shrinkage*, *bshr* for *bottom time series shrinkage*, *hbshr* for *High frequency bottom time series shrinkage*.
$\mathrm{oct}_{o}(\cdot)$	Optimal cross-temporal reconciliation with overlapping residuals (see Section 4.2) for the *wlsv* and *bdshr* approaches (Appendix A).
$\mathrm{oct}_{oh}(\cdot)$	Optimal cross-temporal reconciliation with overlapping and multi-step residuals (see Section 4.1 and 4.2) for the shrinkage approaches presented in Section 4: *shr* stands for *Global shrinkage*, *hshr* for *High frequency shrinkage*.

Table 2: *Cross-temporal reconciliation approaches for the Monte Carlo simulation (see Section 5), the Australian GDP (see Section 6) and the Australian Tourism Demand (see Section 7) forecasting experiments. All the reconciliation procedures are available in the R package* `FoReco` *(Girolimetto & Di Fonzo 2023).*

5.1 Covariance matrix comparison and probabilistic accuracy scores

To compare the true covariance matrix $\boldsymbol{\Omega}_{ct}$ with the estimated covariance matrix $\widehat{\boldsymbol{\Omega}}$, we use the Frobenius norm to quantify the difference between two matrices:

$$\|\boldsymbol{D}\|_F = \sqrt{\sum_{i=1}^{n(k^*+m)} \sum_{j=1}^{n(k^*+m)} |d_{i,j}|^2}$$

where $\boldsymbol{D} = \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_{ct}$. The true covariance matrix, shown in Figure 7, was compared to the estimated covariance matrices obtained using various reconciliation approaches and techniques for generating sample paths of the base forecasts. Thus, we should be able to determine which reconciliation approach and simulation technique produce an accurate estimate of the covariance matrix. Other types of matrix norms were also considered with similar results.

Reconciliation approach	Generation of the base forecasts paths				
	ctjb	Gaussian approach*			
		G	B	H	HB
base	8.260	7.748	6.549	3.409	2.215
ct(bu)	3.195	2.215	2.215	**2.215**	2.215
ct(shr_{cs}, bu_{te})	3.202	2.224	2.215	2.224	2.215
ct$(wlsv_{te}, bu_{cs})$	**3.183**	**2.188**	2.188	**2.215**	2.215
oct$(wlsv)$	3.766	3.082	2.191	2.910	2.215
oct$(bdshr)$	3.203	2.195	**2.184**	2.224	**2.215**
oct$_h(shr)$	3.251	2.260	2.202	2.226	2.215
oct$_h(bshr)$	3.602	2.720	2.220	2.756	2.215
oct$_h(hshr)$	4.869	4.138	4.167	2.225	2.215
oct$_h(hbshr)$	5.731	5.085	4.167	2.756	2.215

*The Gaussian method employs a sample covariance with multi-step residuals.

Table 3: *Frobenius norm between the true (in Figure 7) and the estimated covariance matrix for different reconciliation approaches and different techniques for simulating the base forecasts. Entries in bold represent the lowest value for each column, while the blue entry represent the global minimum. The reconciliation approaches are described in Table 2.*

From Table 3, it appears that the reconciled covariance matrices are always closer to the true matrix than the base forecast matrix when using both the Gaussian and the bootstrap approach. Overall, there are no major differences in the findings when using either one-step or multi-step residuals in cross-temporal forecast reconciliation. In fact, using approaches like oct$(bdshr)$, we obtain results that are consistent with approaches such as oct$_h(shr)$, where no temporal and/or cross-sectional incorrelation assumptions are imposed. It is worth noting that the HB covariance matrix when used to calculate the base forecasts samples, is not changed by the reconciliation step (see Appendix B). In conclusion, our results suggest that using multi-step residuals or bootstrap techniques may help find a "good" estimate of the covariance matrix, which can be further improved by the reconciliation.

The accuracy of the probabilistic forecasts is evaluated using the Continuous Ranked Probability Score (CRPS, Gneiting & Katzfuss 2014), given by

$$\text{CRPS}(\widehat{P}_i, z_i) = \frac{1}{L}\sum_{l=1}^{L}|x_{i,l} - z_i| - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{j=1}^{L}\left|x_{i,l} - x_{i,j}\right|, \quad i = 1, \ldots, n, \tag{15}$$

where $\widehat{P}_i(\omega) = \dfrac{1}{L}\displaystyle\sum_{l=1}^{L}\mathbf{1}\left(x_{i,l} \leq \omega\right)$, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L \in \mathbb{R}^n$ is a collection of L random draws from the predictive distribution and $\boldsymbol{z} \in \mathbb{R}^n$ is the observation vector. CRPS[5] is an index that considers the single series and provides us a marginal evaluation of the approaches. In addition, we employ the Energy Score (ES, Gneiting & Katzfuss 2014), that is the CRPS extension to the multivariate case, to evaluate the forecasting accuracy for the whole system (Panagiotelis et al. 2023, Wickramasuriya 2023),

$$\text{ES}(\widehat{P}, \boldsymbol{z}) = \frac{1}{L}\sum_{l=1}^{L}\|\boldsymbol{x}_l - \boldsymbol{z}\|_2 - \frac{1}{2(L-1)}\sum_{i=1}^{L-1}\|\boldsymbol{x}_l - \boldsymbol{x}_{l+1}\|_2 \tag{16}$$

[5]The CRPS is implemented in the R package `scoringRules` (Jordan et al. 2019)

Reconciliation approach	Generation of the base forecasts sample paths									
	ctjb	Gaussian approach*				ctjb	Gaussian approach*			
		G	B	H	HB		G	B	H	HB
	$\forall k \in \{2,1\}$					$k=1$				
base	1.000	0.998	0.999	1.002	1.004	1.000	0.998	0.999	0.999	1.000
ct(bu)	0.901	0.900	0.900	0.900	0.900	0.978	0.976	0.976	0.977	0.977
ct(shr_{cs}, bu_{te})	**0.901**	**0.900**	**0.899**	**0.900**	**0.900**	**0.977**	**0.976**	**0.976**	**0.976**	**0.976**
ct($wlsv_{te}, bu_{cs}$)	0.910	0.916	0.916	0.916	0.917	0.986	0.993	0.993	0.993	0.993
oct($wlsv$)	0.922	0.930	0.930	0.930	0.931	0.998	1.006	1.006	1.007	1.007
oct($bdshr$)	0.910	0.916	0.915	0.916	0.916	0.986	0.992	0.992	0.993	0.993
oct$_h$(shr)	0.904	0.903	0.902	0.902	0.903	0.980	0.979	0.978	0.979	0.979
oct$_h$($bshr$)	0.923	0.922	0.922	0.921	0.922	0.999	0.998	0.998	0.998	0.998
oct$_h$($hshr$)	0.974	0.972	0.972	0.974	0.975	1.052	1.050	1.050	1.053	1.053
oct$_h$($hbshr$)	0.987	0.985	0.985	0.987	0.989	1.065	1.063	1.064	1.066	1.068
	$k=2$									
base	1.000	0.998	0.999	1.005	1.008					
ct(bu)	0.831	0.830	0.829	0.829	0.830					
ct(shr_{cs}, bu_{te})	**0.830**	**0.830**	**0.829**	**0.829**	**0.830**					
ct($wlsv_{te}, bu_{cs}$)	0.840	0.846	0.844	0.845	0.846					
oct($wlsv$)	0.851	0.859	0.859	0.859	0.861					
oct($bdshr$)	0.839	0.845	0.844	0.845	0.846					
oct$_h$(shr)	0.834	0.833	0.831	0.832	0.832					
oct$_h$($bshr$)	0.852	0.851	0.851	0.851	0.852					
oct$_h$($hshr$)	0.902	0.900	0.899	0.901	0.902					
oct$_h$($hbshr$)	0.915	0.913	0.913	0.914	0.917					

*The Gaussian method employs a sample covariance matrix and includes four techniques (G, B, H, HB) with multi-step residuals.

Table 4: *AvgRelCRPS defined in* (17) *and* (18). *A lower value, indicates a more accurate forecast. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

where $\|\cdot\|_2$ is the L_2 norm.

In Tables 4 and 5 are reported the results using these two scores. In particular, we consider the geometric mean of the relative CRPS (Fleming & Wallace 1986), and the relative ES:

$$\text{AvgRelCRPS}_{j,s}^{[k]} = \left(\prod_{i=1}^{n} \frac{CRPS_{i,j,s}^{[k]}}{CRPS_{i,0,0}^{[k]}}\right)^{\frac{1}{n}} \quad \text{and} \quad \text{RelES}_{j,s}^{[k]} = \frac{ES_{j,s}^{[k]}}{ES_{0,0}^{[k]}} \tag{17}$$

where j denotes the reconciliation approach used and s indicates the approach used to simulate the base forecasts. As a reference approach ($s=0$ and $j=0$), we consider the base forecasts using the Bootstrap approach. Low values indicate better quality of forecasts. If we consider all the temporal aggregation orders (i.e. $\forall k \in \{2,1\}$), we use the geometric averages

$$\text{AvgRelCRPS}_{j,s} = \left(\prod_{\substack{i=1,\dots,n \\ k\in\{1,2\}}} \frac{CRPS_{i,j,s}^{[k]}}{CRPS_{i,0,0}^{[k]}}\right)^{\frac{1}{n(k^*+m)}}, \tag{18}$$

and

$$\text{AvgRelES}_{j,s} = \left(\prod_{k\in\{1,2\}} \frac{ES_{j,s}^{[k]}}{ES_{0,0}^{[k]}}\right)^{\frac{1}{(k^*+m)}}. \tag{19}$$

Reconciliation approach	Generation of the base forecasts sample paths									
	ctjb	Gaussian approach*				ctjb	Gaussian approach*			
		G	B	H	HB		G	B	H	HB
	$\forall k \in \{2,1\}$					$k=1$				
base	1.000	0.996	0.999	1.000	1.004	1.000	0.997	1.000	0.997	1.000
ct(bu)	0.897	0.895	0.896	0.897	**0.895**	0.969	**0.967**	**0.967**	0.968	**0.968**
ct(shr_{cs}, bu_{te})	**0.896**	**0.895**	**0.895**	**0.896**	0.896	**0.968**	0.968	0.967	**0.968**	0.968
ct($wlsv_{te}, bu_{cs}$)	0.906	0.912	0.911	0.910	0.912	0.977	0.984	0.983	0.981	0.984
oct($wlsv$)	0.916	0.923	0.923	0.923	0.924	0.989	0.994	0.995	0.995	0.997
oct($bdshr$)	0.906	0.910	0.910	0.911	0.912	0.977	0.981	0.982	0.983	0.985
oct$_h$(shr)	0.900	0.898	0.898	0.897	0.898	0.971	0.969	0.969	0.969	0.969
oct$_h$($bshr$)	0.916	0.914	0.916	0.915	0.916	0.987	0.986	0.987	0.987	0.988
oct$_h$($hshr$)	0.967	0.964	0.964	0.966	0.967	1.040	1.036	1.036	1.040	1.040
oct$_h$($hbshr$)	0.978	0.975	0.976	0.977	0.980	1.051	1.047	1.049	1.051	1.052
	$k=2$									
base	1.000	0.996	0.998	1.003	1.008					
ct(bu)	0.831	0.829	0.829	0.830	**0.828**					
ct(shr_{cs}, bu_{te})	**0.829**	**0.828**	**0.829**	**0.829**	0.829					
ct($wlsv_{te}, bu_{cs}$)	0.839	0.844	0.844	0.844	0.845					
oct($wlsv$)	0.849	0.858	0.856	0.856	0.857					
oct($bdshr$)	0.839	0.845	0.843	0.845	0.844					
oct$_h$(shr)	0.835	0.833	0.833	0.831	0.832					
oct$_h$($bshr$)	0.850	0.847	0.849	0.849	0.850					
oct$_h$($hshr$)	0.900	0.897	0.896	0.897	0.899					
oct$_h$($hbshr$)	0.910	0.907	0.908	0.909	0.912					

*The Gaussian method employs a sample covariance matrix and includes four techniques (G, B, H, HB) with multi-step residuals.

Table 5: *ES ratio indices defined in* (17) *and* (19). *A lower value, indicates a more accurate forecast. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

A limitation of this simulation setting is that we are using a high number of residuals, which may result in undervaluing the benefit from using the parametrisation form of the covariance matrix such as HB, H, or B (see Section 4). Additionally, shrinkage techniques often yield very similar results when we use the corresponding matrix with $\lambda = 0$ (full covariance matrix). It is important to consider these limitations when interpreting the results.

The good performance of the ct(bu) approach can be explained by a good quality of the base forecasts at the bottom level for $k = 1$, and therefore it is difficult for the other approaches to correctly adjust them using the somewhat less good forecasts of the higher temporal and cross-sectional levels. This also explains the good performance of ct(shr_{cs}, bu_{te}), which by definition only takes into account the information provided by the most temporally disaggregated base forecasts. Looking at the optimal cross-temporal reconciliation approaches, it does not seem to be any advantage in using multi-step residuals to calculate the covariance matrix in the reconciliation step.

In conclusion, studing the use of simulation and reconciliation techniques to improve the accuracy of forecasts, we found that simulating base forecasts from multi-step residuals allows us to estimate a covariance matrix close to the true one. Additionally, we observed that reconciliation could be used to further improve the accuracy of these estimates: accurate base

forecasts for $k = 1$ assist the good performance for bottom-up and optimal cross-temporal reconciliation approaches, such as oct($wlsv$) and oct($bdshr$), which perform well in terms of both CRPS and ES.

6 Forecasting Australian GDP

The Australian Quarterly National Accounts dataset has been widely used in the literature on cross-sectional and cross-temporal reconciliation (Athanasopoulos et al. 2020, Bisaglia et al. 2020, Di Fonzo & Girolimetto 2022*c*, 2023*a*). In particular, Athanasopoulos et al. (2020) proposed using state-of-the-art forecast reconciliation methods to improve the accuracy of macroeconomic forecasts and facilitate aligned decision-making. In their empirical analysis, they applied cross-sectional forecast reconciliation to 95 Australian Quarterly National Accounts time series that represent the Gross Domestic Product (GDP) calculated using both the income and expenditure approaches. These two approaches correspond to two distinct hierarchical structures, with GDP at the top and 15 lower-level aggregates in the income approach, and GDP as the top-level aggregate in a hierarchy of 79 time series in the expenditure approach (for more information, see Athanasopoulos et al. 2020, pp. 702–705 and figures 21.4–21.7). Bisaglia et al. (2020) showed how to obtain a "one-number" forecast where the GDP reconciled forecasts are coherent for both the expenditure and income sides. Di Fonzo & Girolimetto (2022*d*,*c*) extended the one number forecasts idea to obtain fully reconciled probabilistic forecasts, and Di Fonzo & Girolimetto (2023*a*) computed cross-temporally reconciled point forecasts. Building on these results, now we compute the cross-temporally probabilistic forecasts.

We use univariate ARIMA models to obtain quarterly base forecasts for the $n = 95$ separate time series (over the period 1984:Q4 – 2018:Q1), using the `auto.arima` function from the R package `forecast` (Hyndman et al. 2023), and perform a rolling forecast experiment with an expanding window: the first training sample spanned the period 1984:Q4 to 1994:Q3, and the last ended in 2017:Q1, for a total of 91 forecast origins. For the temporal aggregation dimension we aggregate the quarterly data to both semi-annual and annual. We obtain 4-step, 2-step and 1-step ahead base forecasts respectively from the quarterly, semi-annual and annual frequencies, i.e., $\mathcal{K} = \{4, 2, 1\}$.

The base forecast samples in the Gaussian case are obtained using the sample[6] covariance matrices with the *Global* (G) and *High frequency* (H) parametrisation (Section 4), since it is not possible to identify a unique representation for the other cases[7]. We compare the results obtained using multi-step residuals with and without overlapping, in order to measure the benefit of obtaining overlapping residuals. In the nonparametric case, we use the cross-temporal joint bootstrap presented in Section 3.2. Finally, to reconcile the resulting (1000) base

[6]The results with shrunk covariance matrices are available in the online appendix.

[7]When simultaneously considering Income and Expenditure sides hierarchies, the result is a general linearly constrained time series, where bottom and upper time series are not uniquely defined, making unfeasible the cross-sectional bottom-up reconciliation approach(Di Fonzo & Girolimetto 2022*d*).

forecasts samples, we have applied the following techniques[8] (see Table 2): ct(shr_{cs}, bu_{te}), ct(wls_{cs}, bu_{te}), oct$_o(wlsv)$, oct$_o(bdshr)$, oct$_{oh}(shr)$ and oct$_{oh}(hshr)$.

6.1 Results

Forecasting accuracy indices based on CRPS and ES are presented in Tables 6 and 7, respectively. As benchmark approach, we use the base forecasts calculated using the bootstrap method. At the base forecasts level, we find that using a parametric approach with the normal distribution performs better than the non-parametric bootstrap approach. This is likely due to the limited number of residuals available for bootstrapping, which does not allow for sufficient exploration of the data. Regarding reconciliation, diagonal matrices were found to be more effective than matrices that try to recover the correlation structure through shrinkage. Among all the procedures, ct(wls_{cs}, bu_{te}) and oct$_o(wlsv)$ show the greatest relative gains. In contrast, oct$_{oh}(shr)$ and $oct_{oh}(hshr)$ do not show much of an improvement. Furthermore, the greatest improvements are observed for higher temporal aggregation levels.

Figure 8 shows the base (top panels) and reconciled (intermediate and bottom panels) forecasts for the GDP at iteration #14 of the forecasting experiment for the Gaussian and the bootstrap approaches (in left and right columns, respectively). Note that in contrast to the Gaussian case, in the nonparametric case, the medians and means are not always overlapping and centered within the forecast interval, which is asymmetrical, as is the Gaussian case.

Additionally, we utilize the non-parametric Friedman test and the post hoc "Multiple Comparison with the Best" (MCB) Nemenyi test (Koning et al. 2005, Kourentzes & Athanasopoulos 2019, Makridakis et al. 2022) to determine if the forecasting performances of the different techniques are significantly different from one another. Figure 9 presents the MCB using the CRPS. We found that ct(wls_{cs}, bu_{te}) and oct$_o(wlsv)$ are significantly better than the base forecasts at any level of aggregation.

Overall, we find that using overlapping residuals almost always leads to a greater improvement in terms of ES and CRPS, and this is also generally true for reconciliation. Forecasts at the most aggregated level (year) seem to benefit the most from reconciliation, and using one-step overlapping residuals appears to be sufficient to improve forecasts if the generation of the base forecasts sample paths take into account the multi-step structure.

7 Forecasting Australian Tourism Demand

The Australian Tourism Demand dataset (Wickramasuriya et al. 2019) measures the number of nights Australians spent away from home. It includes 228 monthly observations of Visitor Nights (VN) from January 1998 to December 2016, and has a cross-sectional grouped structure based on a geographic hierarchy crossed by purpose of travel. The geographic hierarchy comprises seven states, 27 zones, and 76 regions, for a total of 111 nested geographic divisions. Six of these zones (see Table D.2 in Appendix D) are each formed by a single region, resulting in a total of 105 unique nodes in the hierarchy. The purpose of travel comprises four categories:

[8] In the online appendix, we also report the results obtained using one-step residuals in the reconciliation (thus giving a biased covariance matrix, but with less computational cost).

Reconciliation approach	Generation of the base forecasts sample paths									
	ctjb	Gaussian approach*				ctjb	Gaussian approach*			
		G_h	H_h	G_{oh}	H_{oh}		G_h	H_h	G_{oh}	H_{oh}
	$\forall k \in \{4,2,1\}$					$k=1$				
base	1.000	0.979	0.995	0.968	0.976	1.000	0.988	0.988	0.971	0.971
ct(shr_{cs}, bu_{te})	0.937	0.956	0.956	0.976	0.976	0.992	1.008	1.008	1.029	1.029
ct(wls_{cs}, bu_{te})	0.930	0.917	0.917	0.898	0.898	0.986	0.974	0.975	0.956	0.956
oct$_o$($wlsv$)	**0.926**	**0.911**	**0.912**	**0.896**	**0.895**	**0.984**	**0.971**	**0.970**	**0.954**	**0.954**
oct$_o$($bdshr$)	0.978	0.964	0.946	0.952	0.930	1.034	1.016	1.003	1.005	0.989
oct$_{oh}$(shr)	1.102	1.059	1.001	1.094	0.988	1.172	1.109	1.066	1.160	1.059
oct$_{oh}$($hshr$)	1.006	0.983	1.009	0.974	1.001	1.068	1.046	1.059	1.034	1.061
	$k=2$					$k=4$				
base	1.000	0.984	0.993	0.968	0.976	1.000	0.966	1.004	0.964	0.981
ct(shr_{cs}, bu_{te})	0.949	0.966	0.966	0.987	0.987	0.874	0.896	0.896	0.914	0.914
ct(wls_{cs}, bu_{te})	0.942	0.928	0.928	0.909	0.909	0.866	0.853	0.853	0.834	0.834
oct$_o$($wlsv$)	**0.938**	**0.921**	**0.923**	**0.907**	**0.906**	**0.860**	**0.847**	**0.848**	**0.832**	**0.830**
oct$_o$($bdshr$)	0.991	0.974	0.957	0.964	0.942	0.914	0.905	0.883	0.892	0.865
oct$_{oh}$(shr)	1.120	1.069	1.013	1.113	1.002	1.020	1.002	0.928	1.015	0.909
oct$_{oh}$($hshr$)	1.021	0.996	1.021	0.987	1.016	0.934	0.912	0.951	0.904	0.931

*The Gaussian method employs a sample covariance matrix:
G_h and H_h use multi-step residuals and G_{oh} and H_{oh} use overlapping and multi-step residuals.

Table 6: *AvgRelCRPS defined in* (17) *and* (18) *for the Australian Quarterly National Accounts dataset. A lower value, indicates a more accurate forecast. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

Reconciliation approach	Generation of the base forecasts sample paths									
	ctjb	Gaussian approach*				ctjb	Gaussian approach*			
		G_h	H_h	G_{oh}	H_{oh}		G_h	H_h	G_{oh}	H_{oh}
	$\forall k \in \{4,2,1\}$					$k=1$				
base	1.000	0.970	0.988	0.960	0.970	1.000	0.977	0.977	0.965	0.965
ct(shr_{cs}, bu_{te})	0.897	0.944	0.944	0.973	0.973	0.964	1.001	1.001	1.033	1.033
ct(wls_{cs}, bu_{te})	**0.886**	0.880	**0.880**	**0.860**	**0.860**	**0.954**	**0.944**	**0.945**	**0.928**	**0.928**
oct$_o$($wlsv$)	0.891	**0.879**	0.881	0.864	0.864	0.958	0.945	0.945	0.931	0.931
oct$_o$($bdshr$)	0.940	0.928	0.910	0.918	0.895	1.004	0.986	0.971	0.980	0.961
oct$_{oh}$(shr)	1.059	1.015	0.956	1.053	0.945	1.130	1.063	1.019	1.121	1.016
oct$_{oh}$($hshr$)	0.986	0.968	0.999	0.959	0.992	1.053	1.034	1.049	1.024	1.055
	$k=2$					$k=4$				
base	1.000	0.972	0.985	0.959	0.969	1.000	0.959	1.000	0.957	0.976
ct(shr_{cs}, bu_{te})	0.915	0.961	0.960	0.991	0.991	0.818	0.874	0.874	0.899	0.900
ct(wls_{cs}, bu_{te})	**0.904**	0.896	**0.896**	**0.877**	**0.877**	**0.807**	0.805	**0.805**	**0.782**	**0.783**
oct$_o$($wlsv$)	0.908	**0.895**	0.898	0.881	0.882	0.812	**0.802**	0.806	0.786	0.786
oct$_o$($bdshr$)	0.960	0.947	0.929	0.938	0.915	0.860	0.856	0.836	0.841	0.816
oct$_{oh}$(shr)	1.082	1.029	0.973	1.076	0.963	0.971	0.954	0.882	0.967	0.861
oct$_{oh}$($hshr$)	1.007	0.988	1.017	0.979	1.014	0.904	0.888	0.934	0.881	0.913

*The Gaussian method employs a sample covariance matrix:
G_h and H_h use multi-step residuals and G_{oh} and H_{oh} use overlapping and multi-step residuals.

Table 7: *ES ratio indices defined in* (17) *and* (19) *for the Australian Quarterly National Accounts dataset. A lower value, indicates a more accurate forecast. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

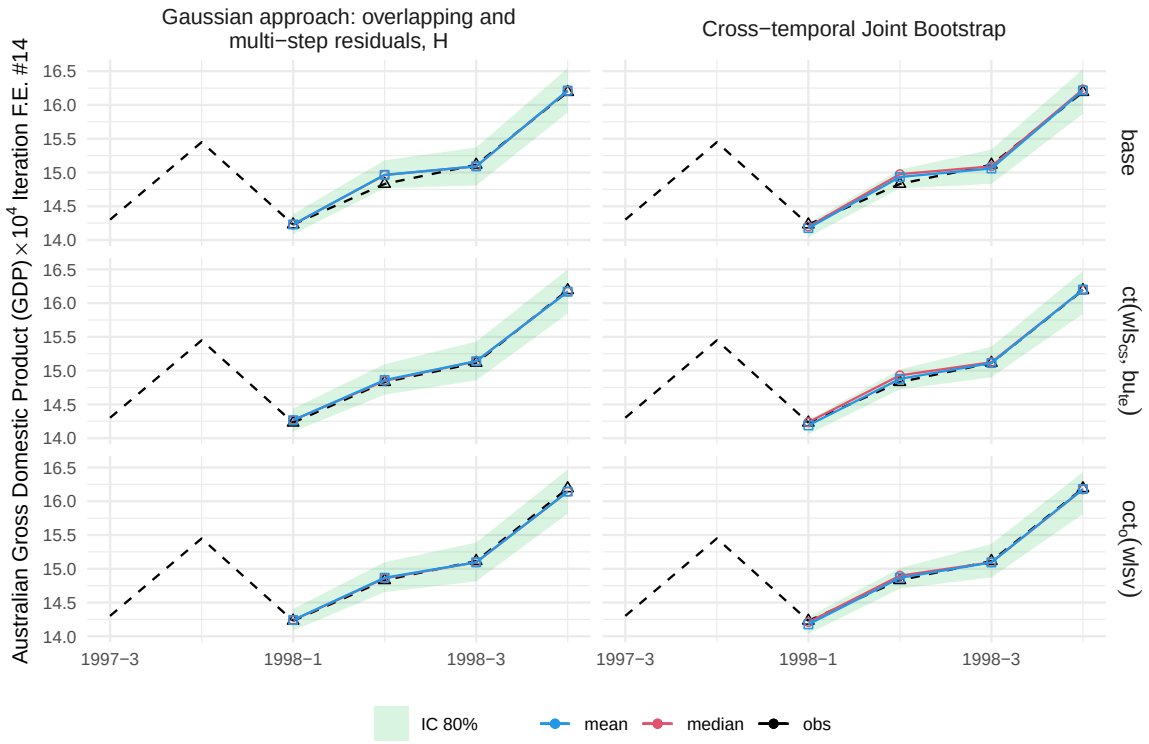

Figure 8: *Base (top panels) and reconciled (intermediate and bottom panels) forecasts for Australian GDP at iteration* #14 *of our forecasting experiment for the Gaussian (left) and bootstrap (right) approaches. Green shaded areas represent the 80% forecast interval. Black lines and triangles indicate the observed true values, the red line represent the median value, and the blue line represent the mean value.*

holiday (Hol), visiting friends and relatives (VFR), business (Bus), and other (Oth). To avoid redundancies (Di Fonzo & Girolimetto 2022*b*), 24 nodes are not considered, resulting in an unbalanced hierarchy of 525 unique nodes instead of the theoretical 555 with duplicated nodes. The dataset includes the 304 bottom series, which are aggregated into 221 upper time series. Table 8 omits duplicated entries and updates the information in Table 7 from Wickramasuriya et al. (2019). This data can be temporally aggregated into 2, 3, 4, 6, or 12 months ($\mathcal{K} = \{12, 4, 3, 2, 1\}$).

The forecasting experiment involves using a recursive training sample with an expanding window. The process begins by using the first 10 years of data, from January 1998 to December 2008, to generate forecasts for the entire following year (2009). Then, the training set is increased by one month. This process is repeated until the last training set is used (January 1998 to December 2015) with a total of 85 different forecast origins. For the temporal aggregation dimension we aggregate the monthly data up to annual data. We obtain 12-step, 6-step, 4-step, 3-step, 2-step and 1-step ahead base forecasts respectively from the monthly data and the aggregation over 2, 3, 4, 6, and 12 months. ETS models selected by minimizing the AICc criterion with the R package `forecast` are fitted to the log-transformed data, with the resulting base forecasts being back-transformed to produce non-negative forecasts, as described in Wickramasuriya et al. (2020).

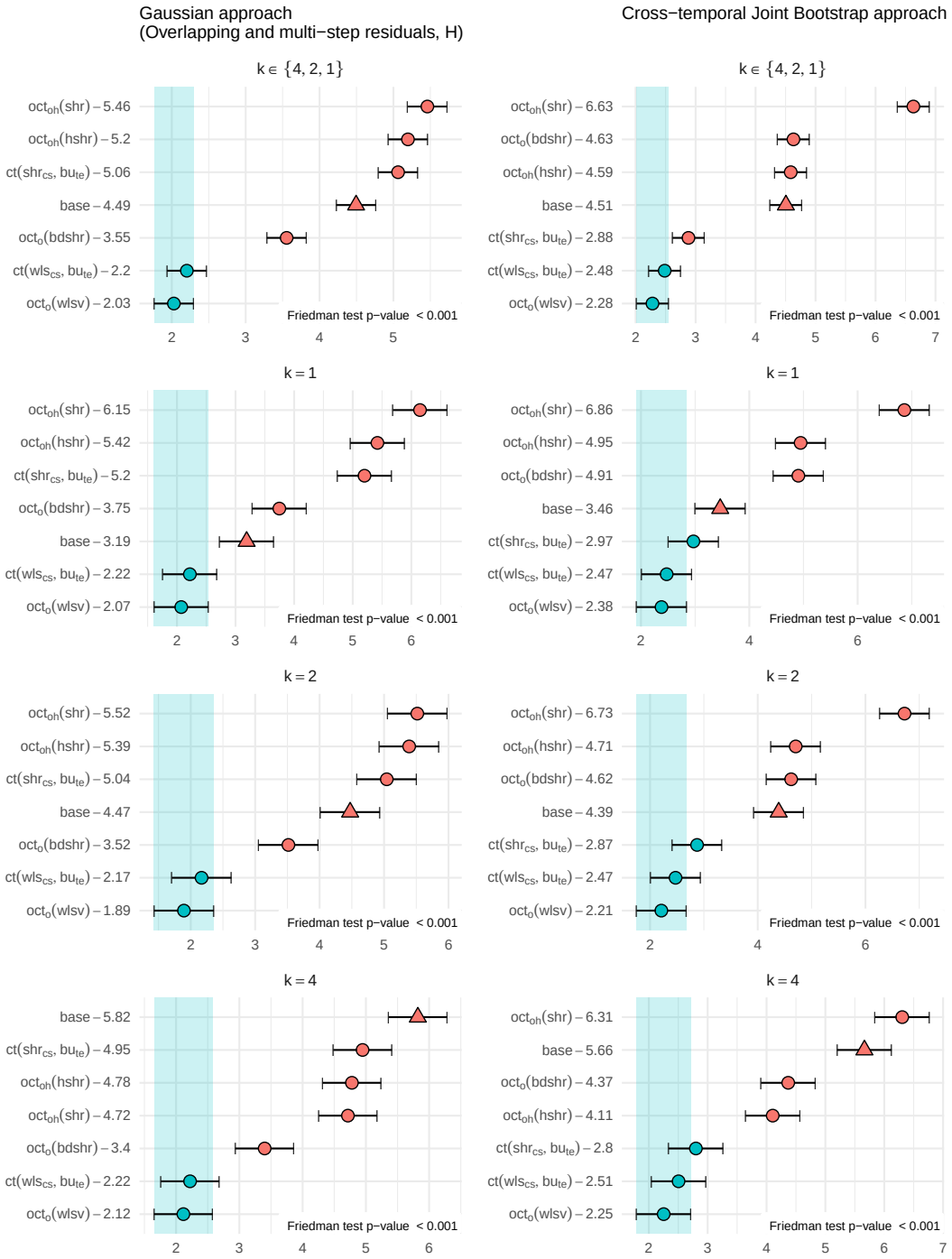

Figure 9: *"Multiple Comparison with the Best" (MCB) Nemenyi test using CRPS at different temporal aggregation levels for the Gaussian (using overlapping and multi-step residuals, H) and the non-parametric bootstrap approaches. In each panel, the Friedman test p-value is reported in the lower right corner. The mean rank of each approach is shown to the right of its name. Statistical differences in performance are indicated if the intervals of two forecast reconciliation procedures do not overlap. Thus, approaches that do not overlap with the blue interval are considered significantly worse than the best, and vice-versa.*

	Number of series		
	GD	**PT**	Tot.
Australia	1	4	5
States	7	28	35
Zones*	21	84	105
Regions	76	304	380
Total	**105**	**420**	**525**

* 6 Zones with only one Region are included in Regions.

Table 8: *Grouped time series for Australian tourism flows. GD: Geographic Division; PT: Purpose of Travel.*

The (1000) base forecast samples are obtain using the Gaussian approach with sample[9] covariance matrices (Section 4) using multi-step residuals[10] and the bootstrap approach Section 3.2. For reconciliation, 11 different approaches have been adopted (see Table 2): ct(bu), ct(shr_{cs}, bu_{te}), ct$(wlsv_{te}, bu_{cs})$, ct(ols), oct$(struc)$, oct$(wlsv)$, oct$(bdshr)$, oct$_h(hbshr)$, oct$_h(bshr)$, oct$_h(hshr)$, and oct$_h(shr)$.

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. For the base forecasts, using the bootstrap approach produces forecasts naturally non negative (ETS model with the log-transformation), while this is not true for the Gaussian approach. In this case, any negative forecast is set equal to zero. For the cross-temporal reconciliation, Di Fonzo & Girolimetto (2023*a*) propose two solutions: either a state-of-the-art numerical optimization procedure (`osqp`, Stellato et al. 2022, 2020), or a simple heuristic strategy called set-negative-to-zero (sntz). With sntz, any negative high frequency bottom time series reconciled forecasts are set to zero, and then a cross-temporal reconciliation bottom-up (see Section 2.2) is used to obtain the complete set of fully coherent forecasts. Di Fonzo & Girolimetto (2023*b*) found that both methods produce similar quality forecasts, but the optimization method required much more time and computation effort compared to the sntz heuristic. To reduce computational demands, we chose to use the less time-intensive heuristic approach for reconciliation. Following Theorem 3.1, we are able to obtain the reconciled sample respecting non-negativity constraints starting from an incoherent sample simulated from a Gaussian distribution.

Finally, to evaluate the performance, we employed the Continuous Ranked Probability Score (CRPS), the Energy Score (ES), and the "Multiple Comparison with the Best" (MCB) Nemenyi test, introduced and discussed in Sections 5.1 and 6.1.

7.1 Results

The CRPS and ES indices are shown, respectively, in Tables 9 and 10 for monthly, quarterly and annual forecasts[11]. These tables are divided by different temporal levels and each column

[9] The results with shrunk covariance matrices are available in the online appendix.

[10] We do not include overlapping, as we are unable to correctly determine the residuals for the overlapping series using ETS models (see Section 4.2).

[11] The complete set of results for all temporal aggregation levels is reported in the online appendix.

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	0.971	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**
oct$_h$($hbshr$)	0.989	1.018	1.020	1.016	1.018	0.982	1.004	1.007	1.004	1.009
oct$_h$($bshr$)	0.994	1.018	1.020	1.016	1.019	0.988	1.007	1.013	1.006	1.012
oct$_h$($hshr$)	**0.969**	0.993	0.993	0.990	0.991	0.953	0.977	0.977	0.979	0.979
oct$_h$(shr)	1.007	0.980	0.972	0.970	0.970	1.000	0.986	0.977	0.976	0.974
		$k=3$					$k=12$			
base	1.000	0.971	0.971	0.972	0.973	1.000	0.968	0.967	0.969	0.969
ct(bu)	1.273	1.010	1.010	1.010	1.010	1.675	1.038	1.037	1.037	1.038
ct(shr_{cs}, bu_{te})	1.041	0.977	0.974	0.977	0.974	1.163	0.977	0.965	0.977	0.965
ct($wlsv_{te}, bu_{cs}$)	1.046	0.976	0.976	0.974	0.974	1.174	0.978	0.978	0.971	0.971
oct(ols)	0.994	0.992	0.993	0.991	0.992	0.982	0.982	0.983	0.980	0.975
oct($struc$)	0.986	0.967	0.966	0.966	0.965	0.982	0.951	0.949	0.947	0.943
oct($wlsv$)	0.983	0.963	0.962	0.962	0.962	1.025	0.954	0.953	0.949	0.947
oct($bdshr$)	0.972	**0.960**	**0.958**	**0.957**	**0.957**	1.002	**0.950**	**0.944**	**0.939**	**0.935**
oct$_h$($hbshr$)	0.994	1.019	1.021	1.018	1.020	0.982	1.027	1.029	1.024	1.021
oct$_h$($bshr$)	0.999	1.021	1.022	1.018	1.022	0.987	1.024	1.021	1.021	1.019
oct$_h$($hshr$)	**0.971**	0.994	0.994	0.992	0.993	**0.978**	1.003	1.005	0.996	0.997
oct$_h$(shr)	1.009	0.986	0.978	0.976	0.976	1.010	0.963	0.956	0.952	0.952

*The Gaussian method employs a sample covariance matrix and includes four techniques (G, B, H, HB) with multi-step residuals..

Table 9: *AvgRelCRPS defined in* (17) *and* (18) *for Australian Tourism Demand. A lower value, indicates a more accurate forecast. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

uses a different approach to calculate the base forecasts, referred to as "base". The bootstrap method was used as a benchmark to calculate the accuracy indices.

Base forecasts using a Gaussian approach are better in terms of both CRPS and ES compared to the bootstrap approach (the benchmark). Assumptions of truncated Gaussianity (Gaussian with negative values set to zero) may seem strict, but given the limited number of residuals, it can lead to improved forecasts in terms of CRPS and ES. Bootstrap forecasts suffer from the limited number of available residuals, leading in general to lower predictive accuracy. The Gaussian approach overcomes this limitation and provides better results. Regarding the different covariance matrix estimates for Gaussian base forecasts, there are no big differences. For this reason, using only the high frequency bottom time series can be useful to estimate fewer parameters and reduce the initial high dimensionality.

In the Gaussian case, bottom-up ct(bu) and partly bottom-up techniques like ct(shr_{cs}, bu_{te}) and ct($wlsv_{te}, bu_{cs}$) lead to better results than the benchmark (bootstrap base forecasts). However, it's not always guaranteed that the improvement is higher than the starting base forecasts (by comparing the value of each column). This is particularly true for higher levels of temporal aggregation ($k \in \{12,6,4,3\}$, see online appendix for details). Overall,

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
oct$_h$($hbshr$)	0.956	0.889	0.886	0.888	0.884	0.975	0.937	0.936	0.937	0.935
oct$_h$($bshr$)	**0.931**	**0.867**	**0.866**	**0.863**	**0.860**	**0.965**	**0.927**	0.927	0.925	0.923
oct$_h$($hshr$)	1.081	0.935	0.931	0.935	0.927	1.028	0.952	0.951	0.952	0.950
oct$_h$(shr)	1.068	0.899	0.878	0.875	0.864	1.023	0.935	**0.923**	**0.921**	**0.916**
		$k=3$					$k=12$			
base	1.000	0.961	0.958	0.960	0.955	1.000	0.942	0.947	0.951	0.937
ct(bu)	2.428	0.998	0.997	0.997	0.997	2.990	0.922	0.921	0.923	0.923
ct(shr_{cs}, bu_{te})	1.245	0.911	0.904	0.911	0.904	1.326	0.779	0.767	0.777	0.766
ct($wlsv_{te}, bu_{cs}$)	1.500	0.991	0.991	0.986	0.987	1.679	0.917	0.917	0.906	0.908
oct(ols)	0.976	0.918	0.915	0.917	0.912	0.872	0.783	0.784	0.783	0.779
oct($struc$)	1.096	0.939	0.936	0.938	0.933	1.077	0.826	0.822	0.823	0.818
oct($wlsv$)	1.142	0.953	0.949	0.951	0.946	1.149	0.851	0.845	0.847	0.840
oct($bdshr$)	1.060	0.926	0.920	0.921	0.915	1.021	0.808	0.796	0.796	0.787
oct$_h$($hbshr$)	0.975	0.915	0.912	0.915	0.909	0.872	0.775	0.772	0.772	0.770
oct$_h$($bshr$)	**0.954**	**0.895**	**0.895**	**0.892**	**0.887**	**0.833**	**0.741**	**0.741**	**0.737**	**0.735**
oct$_h$($hshr$)	1.093	0.955	0.951	0.956	0.949	1.066	0.851	0.846	0.848	0.838
oct$_h$(shr)	1.082	0.923	0.903	0.900	0.890	1.043	0.797	0.768	0.764	0.750

*The Gaussian method employs a sample covariance matrix and includes four techniques (G, B, H, HB) with multi-step residuals..

Table 10: *ES ratio indices defined in* (17) *and* (19) *for Australian Tourism Demand. A lower value, indicates a more accurate forecast. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

oct($bdshr$) in terms of CRPS is almost always the best. The shrinkage approach oct$_h$($hshr$) has good performance in the bootstrap case: it is competitive with oct($bdshr$) at lower temporal frequency ($k \in \{2,1\}$, see online appendix for details) and it is able to improve for $k \geq 3$. In terms of ES, oct($bdshr$) is still competitive, although it does not always show the best relative performance. In this case, approaches that attempt to establish a temporal and cross-sectional relationship, such as oct$_h$($hbshr$) and oct$_h$($bshr$), are better. It is also worth noting that techniques that don't make use of residuals like oct(ols) and oct($struc$) prove to be competitive by consistently improving the base forecasts in terms of both CRPS and ES.

Figures 10 shows the MCB using the CRPS for the Gaussian approach using multi-step residuals (HB) and the non-parametric bootstrap approach. In general, partly bottom-up procedure improves with respect to base forecasts at monthly level, but optimal cross-temporal procedures are always better. In the bootstrap framework, we can identify a group of three procedures, oct($bdshr$), oct($hshr$) and oct($struc$) that are almost always in the group of the best approaches (denoted by the blue dot). In the Gaussian framework, oct($wlsv$), oct($struc$), and oct($bdshr$) are always significantly better than base forecasts and equivalent in

terms of results for temporal aggregation orders greater than 2. For monthly series, oct$(bdshr)$ is always significantly better than all other approaches.

8 Conclusion

In this paper, we extend the probabilistic reconciliation results setting developed by Panagiotelis et al. (2023) for the cross-sectional case to the cross-temporal framework. Through appropriate notation, we show how theorems and definitions valid for the cross-sectional case can be reinterpreted and extended. The general notation proposed can help investigate extensions following different probabilistic approaches, such as those in Jeon et al. (2019), Ben Taieb et al. (2021) and Corani et al. (2022). We propose a Gaussian and a bootstrap approach to simulate the base forecasts able to capture both temporal and cross-temporal relationships simultaneously, opening the way for further research into cross-temporal probabilistic forecasting.

Moreover, we analyze the use of residuals, showing that one-step residuals fail to capture the temporal structure, and we propose multi-step residuals that can fully capture the full cross-temporal relationships. When dealing with covariance matrices (due to the high-dimensionality of the cross-temporal setting), we propose four alternative forms to reduce the number of parameters to be estimated, showing that the overlapping residuals may reduce the high-dimensionality burden by increasing the number of available residuals. We present a simple simulation to investigate using the different types of residuals in the cross-temporal setting arriving to some useful conclusions. These ideas are worth requiring further investigation in future works.

Finally, we perform empirical applications on two Australian data sets commonly used in forecast reconciliation research: GDP from Income and Expenditure sides and Australian Tourism Demand. We find that in both cases optimal cross-temporal reconciliation approaches significantly improve on base forecasts. We also compare these with partly bottom-up techniques that use uni-dimensional reconciliations (either cross-sectional or temporal) and find confirmation that simultaneously exploiting both dimensions in reconciliation gives better results, especially at higher levels of temporal aggregation. In conclusion, reconciliation approaches can play an important role to improve the accuracy of forecasts in a probabilistic framework while achieving the important attribute of producing coherent forecasts.

Acknowledgements

Tommaso Di Fonzo and Daniele Girolimetto acknowledge financial support from project PRIN2017 "HiDEA: Advanced Econometrics for High-frequency Data", 2017RSMPZZ. Rob Hyndman acknowledges the support of the Australian Government through the Australian Research Council Industrial Transformation Training Centre in Optimisation Technologies, Integrated Methodologies, and Applications (OPTIMA), Project ID IC200100009.

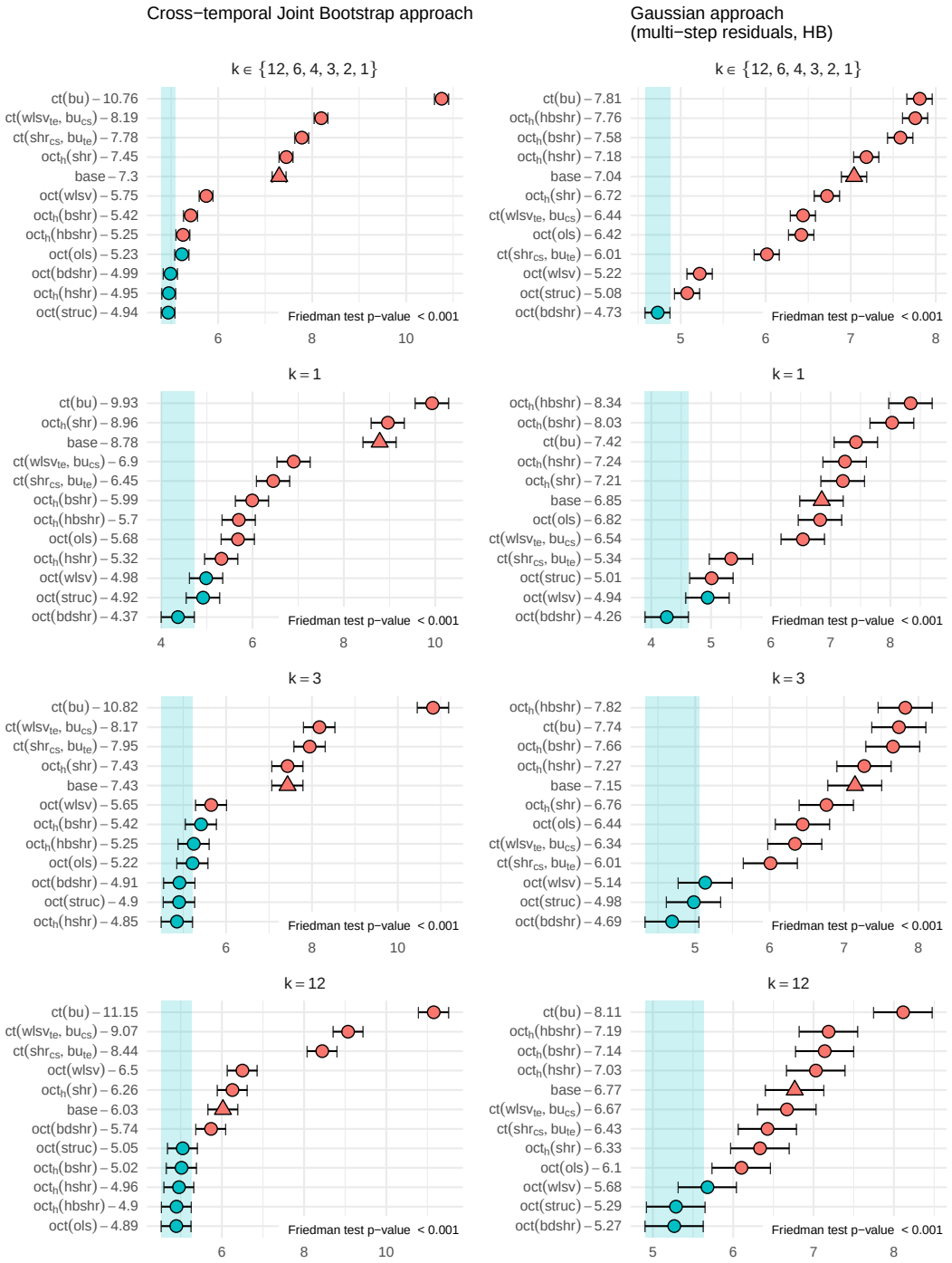

Figure 10: *"Multiple Comparison with the Best" (MCB) Nemenyi test using CRPS at different temporal aggregation orders for the Gaussian (multi-step residuals, HB) and the non-parametric bootstrap approaches. In each panel, the Friedman test p-value is reported in the lower right corner. The mean rank of each approach is shown to the right of its name. Statistical differences in performance are indicated if the intervals of two forecast reconciliation procedures do not overlap. Thus, approaches that do not overlap with the blue interval are considered significantly worse than the best, and vice-versa.*

References

Athanasopoulos, G., Ahmed, R. A. & Hyndman, R. J. (2009), 'Hierarchical forecasts for Australian domestic tourism', *International Journal of Forecasting* **25**(1), 146–166.

Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J. & Affan, M. (2020), Hierarchical Forecasting, *in* P. Fuleky, ed., 'Macroeconomic Forecasting in the Era of Big Data', Vol. 52, Springer International Publishing, Cham, pp. 689–719.

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

Ben Taieb, S., Taylor, J. W. & Hyndman, R. J. (2017), Coherent Probabilistic Forecasts for Hierarchical Time Series, *in* 'Proceedings of the 34th International Conference on Machine Learning', PMLR, pp. 3348–3357.

Ben Taieb, S., Taylor, J. W. & Hyndman, R. J. (2021), 'Hierarchical Probabilistic Forecasting of Electricity Demand With Smart Meter Data', *Journal of the American Statistical Association* **116**(533), 27–43.

Bisaglia, L., Fonzo, T. D. & Girolimetto, D. (2020), Fully reconciled GDP forecasts from Income and Expenditure sides, *in* A. Pollice, N. Salvati & F. Schirripa Spagnolo, eds, 'Book of Short Papers SIS 2020', Pearson, pp. 951–956.

Corani, G., Azzimonti, D., Augusto, J. P. S. C. & Zaffalon, M. (2021), 'Probabilistic Reconciliation of Hierarchical Forecast via Bayes' Rule', *Machine Learning and Knowledge Discovery in Databases* **12459**, 211–226.

Corani, G., Rubattu, N., Azzimonti, D. & Antonucci, A. (2022), 'Probabilistic Reconciliation of Count Time Series', *arXiv:2207.09322* .

Dangerfield, B. J. & Morris, J. S. (1992), 'Top-down or bottom-up: Aggregate versus disaggregate extrapolations', *International Journal of Forecasting* **8**(2), 233–241.

Di Fonzo, T. & Girolimetto, D. (2022*a*), 'Enhancements in cross-temporal forecast reconciliation, with an application to solar irradiance forecasts', *arXiv:2209.07146* .

Di Fonzo, T. & Girolimetto, D. (2022*b*), 'Forecast combination-based forecast reconciliation: Insights and extensions', *International Journal of Forecasting* . in press.

Di Fonzo, T. & Girolimetto, D. (2022*c*), Fully reconciled probabilistic GDP forecasts from Income and Expenditure sides, *in* A. Balzanella, M. Bini, C. Cavicchia & R. Verde, eds, 'Book of Short Papers SIS 2022', Pearson, pp. 1376–1381.

Di Fonzo, T. & Girolimetto, D. (2022*d*), 'Point and probabilistic forecast reconciliation for general linearly constrained multiple time series', *Working paper* .

Di Fonzo, T. & Girolimetto, D. (2023*a*), 'Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives', *International Journal of Forecasting* **39**(1), 39–57.

Di Fonzo, T. & Girolimetto, D. (2023*b*), 'Spatio-temporal reconciliation of solar forecasts', *Solar Energy* **251**, 13–29.

Dunn, D. M., Williams, W. H. & Dechaine, T. L. (1976), 'Aggregate versus Subaggregate Models in Local Area Forecasting', *Journal of the American Statistical Association* **71**(353), 68–71.

Eckert, F., Hyndman, R. J. & Panagiotelis, A. (2021), 'Forecasting Swiss exports using Bayesian forecast reconciliation', *European Journal of Operational Research* **291**(2), 693–710.

Efron, B. (1975), 'Biased versus unbiased estimation', *Advances in Mathematics* **16**(3), 259–277.

Efron, B. & Morris, C. (1975), 'Data Analysis Using Stein's Estimator and its Generalizations', *Journal of the American Statistical Association* **70**(350), 311–319.

Efron, B. & Morris, C. (1977), 'Stein's Paradox in Statistics', *Scientific American* **236**(5), 119–127.

Fleming, P. J. & Wallace, J. J. (1986), 'How not to lie with statistics: The correct way to summarize benchmark results', *Communications of the ACM* **29**(3), 218–221.

Fliedner, G. (2001), 'Hierarchical forecasting: Issues and use guidelines', *Industrial Management & Data Systems* **101**(1), 5–12.

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

Gneiting, T. & Katzfuss, M. (2014), 'Probabilistic Forecasting', *Annual Review of Statistics and Its Application* **1**(1), 125–151.

Gross, C. W. & Sohl, J. E. (1990), 'Disaggregation methods to expedite product line forecasting', *Journal of Forecasting* **9**(3), 233–254.

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.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., Kuroptev, K., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., Garza, F., Girolimetto, D., Ihaka, R., R Core Team, Reid, D., Shaub, D., Tang, Y., Wang, X. & Zhou, Z. (2023), *forecast: Forecasting Functions for Time Series and Linear Models*. R package v8.20. **URL:** *https://pkg.robjhyndman.com/forecast/*

Hyndman, R. J. & Khandakar, Y. (2008), 'Automatic Time Series Forecasting: The forecast Package for R', *Journal of Statistical Software* **27**, 1–22.

Hyndman, R. J., Lee, A. J. & Wang, E. (2016), 'Fast computation of reconciled forecasts for hierarchical and grouped time series', *Computational Statistics & Data Analysis* **97**, 16–32.

Jeon, J., Panagiotelis, A. & Petropoulos, F. (2019), 'Probabilistic forecast reconciliation with applications to wind power and electric load', *European Journal of Operational Research* **279**(2), 364–379.

Jordan, A., Krüger, F. & Lerch, S. (2019), 'Evaluating Probabilistic Forecasts with scoringRules', *Journal of Statistical Software* **90**(12).

Koning, A. J., Franses, P. H., Hibon, M. & Stekler, H. (2005), 'The M3 competition: Statistical tests of the results', *International Journal of Forecasting* **21**(3), 397–409.

Kourentzes, N. & Athanasopoulos, G. (2019), 'Cross-temporal coherent forecasts for Australian tourism', *Annals of Tourism Research* **75**, 393–409.

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

Magnus, J. R. & Neudecker, H. (2019), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edn, Wiley, Hoboken, NJ.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2022), 'M5 accuracy competition: Results, findings, and conclusions', *International Journal of Forecasting* **38**(4), 1346–1364.

Nystrup, P., Lindström, E., Pinson, P. & Madsen, H. (2020), 'Temporal hierarchies with autocorrelation for load forecasting', *European Journal of Operational Research* **280**(3), 876–888.

O'Hara-Wild, M., Hyndman, R. J., Wang, E. & Godahewa, R. (2022), *tsibbledata: Diverse Datasets for 'tsibble'*. R package v0.4.1.
URL: *tsibbledata.tidyverts.org*

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

Panamtash, H. & Zhou, Q. (2018), Coherent Probabilistic Solar Power Forecasting, *in* '2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)', Boise, ID, USA, pp. 1–6.

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., Burkardt, J., Zou, C., Lakiotaki, K. & Chatzipantsiou, C. (2022), *Rfast: A*

Collection of Efficient and Extremely Fast R Functions. R package version 2.0.6.
URL: *https://CRAN.R-project.org/package=Rfast*

Pennings, C. L. & van Dalen, J. (2017), 'Integrated hierarchical forecasting', *European Journal of Operational Research* **263**(2), 412–418.

Punia, S., Singh, S. P. & Madaan, J. K. (2020), 'A cross-temporal hierarchical framework and deep learning for supply chain forecasting', *Computers & Industrial Engineering* **149**, 106796.

R Core Team (2022), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *www.R-project.org*

Sanguri, K., Shankar, S., Punia, S. & Patra, S. (2022), 'Hierarchical container throughput forecasting: The value of coherent forecasts in the management of ports operations', *Computers & Industrial Engineering* **173**, 108651.

Schäfer, J. & Strimmer, K. (2005), 'A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics', *Statistical Applications in Genetics and Molecular Biology* **4**(1).

Spiliotis, E., Petropoulos, F., Kourentzes, N. & Assimakopoulos, V. (2020), 'Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption', *Applied Energy* **261**, 114339.

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

Stellato, B., Banjac, G., Goulart, P., Boyd, S. & Bansal, V. (2022), *osqp: Quadratic Programming Solver using the 'OSQP' Library*. R package v0.6.0.7.
URL: *osqp.org*

Venables, W. N. & Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edn, Springer, New York. ISBN 0-387-95457-0.
URL: *https://www.stats.ox.ac.uk/pub/MASS4/*

Wickramasuriya, S. L. (2023), 'Probabilistic forecast reconciliation under the gaussian framework', *Journal of Business & Economic Statistics, in press* .

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.

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

Yagli, G. M., Yang, D. & Srinivasan, D. (2019), 'Reconciling solar forecasts: Sequential reconciliation', *Solar Energy* **179**, 391–397.

Zambon, L., Azzimonti, D. & Corani, G. (2022), 'Efficient probabilistic reconciliation of forecasts for real-valued and count time series', *arXiv:2210.02286* .

A Cross-sectional, temporal and cross-temporal covariance approximations

Table A.1 presents some approximations for the cross-sectional and the temporal covariance matrices. Nystrup et al. (2020) and Di Fonzo & Girolimetto (2023*a*) provide further temporal reconciliation approaches that exploit possible information in the residual autocorrelation.

Table A.1: *Approximations for the cross-sectional (Hyndman et al. 2011, Hyndman et al. 2016, Wickramasuriya et al. 2019, Di Fonzo & Girolimetto 2023*a*) and temporal (Athanasopoulos et al. 2017, Di Fonzo & Girolimetto 2023*a*) covariance matrix, respectively* $\boldsymbol{W}$ *and* $\boldsymbol{\Omega}$*.*

	Cross-sectional framework	**Temporal framework**
identity	cs(ols): $\boldsymbol{W} = \boldsymbol{I}_n$	te(ols): $\boldsymbol{\Omega} = \boldsymbol{I}_{k^*+m}$
structural	cs($struc$): $\boldsymbol{W} = \text{diag}(\boldsymbol{S}_{cs}\boldsymbol{1}_{nb})$	te($struc$): $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{S}_{te}\boldsymbol{1}_{m})$
series variance	cs(wls): $\boldsymbol{W} = \widehat{\boldsymbol{W}}_D = \boldsymbol{I}_n \odot \widehat{\boldsymbol{W}}$	te($wlsv$): $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Omega}}_{wlsv}$
MinT-shr	cs(shr): $\boldsymbol{W} = \hat{\lambda}\widehat{\boldsymbol{W}}_D + (1-\hat{\lambda})\widehat{\boldsymbol{W}}$	te(shr): $\boldsymbol{\Omega} = \hat{\lambda}\widehat{\boldsymbol{\Omega}}_D + (1-\hat{\lambda})\widehat{\boldsymbol{\Omega}}$
MinT-sam	cs(sam): $\boldsymbol{W} = \widehat{\boldsymbol{W}}$	te(sam): $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Omega}}$

Note: $\widehat{\boldsymbol{W}}$ ($\widehat{\boldsymbol{\Omega}}$) is the covariance matrix of the cross-sectional (temporal) one-step ahead in-sample forecast errors, $\widehat{\boldsymbol{\Omega}}_{wlsv}$ is a diagonal matrix presented by Athanasopoulos et al. (2017), and $\widehat{\boldsymbol{\Omega}}_D = \boldsymbol{I}_{k^*+m} \odot \widehat{\boldsymbol{\Omega}}$, where $\odot$ denotes the Hadamard product.

Di Fonzo & Girolimetto (2023*a*) consider the following approximations for the cross-temporal covariance matrix.

oct(ols) - identity: $\boldsymbol{\Omega}_{ct} = \boldsymbol{I}_{n(k^*+m)}$.

oct($struc$) - structural: $\boldsymbol{\Omega}_{ct} = \text{diag}(\boldsymbol{S}_{ct}\boldsymbol{1}_{mn_b})$.

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) in the temporal framework.

oct($bdshr$) - block-diagonal shrunk cross-covariance scaling: $\boldsymbol{\Omega}_{ct} = \boldsymbol{P}\widehat{\boldsymbol{W}}^{BD}_{ct,shr}\boldsymbol{P}'$, with

$$\widehat{\boldsymbol{W}}^{BD}_{ct,shr} = \begin{bmatrix} \widehat{\boldsymbol{W}}^{[m]}_{shr} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{\frac{m}{k_{p-1}}} \otimes \widehat{\boldsymbol{W}}^{[k_{p-1}]}_{shr} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{I}_m \otimes \widehat{\boldsymbol{W}}^{[1]}_{shr} \end{bmatrix},$$

a block diagonal matrix where $\widehat{\boldsymbol{W}}^{[k]}_{shr}$ is the shrunk estimates of the cross-sectional covariance matrix proposed by Wickramasuriya et al. (2019) for a given temporal aggregation level k and $\boldsymbol{P}$ is the commutation matrix such that $\boldsymbol{P}\text{vec}(\boldsymbol{Y}_\tau) = \text{vec}(\boldsymbol{Y}'_\tau)$.

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 & 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.

oct(sam) - MinT-sam: $\boldsymbol{\Omega}_{ct} = \widehat{\boldsymbol{\Omega}}_{ct}$.

B Alternative forms of the cross-temporal covariance matrix

In this appendix, some derivations of the solutions proposed in Section 4 to obtain an estimator of the cross-temporal covariance matrix are reported. Starting from the the definition of cross-temporal covariance matrix we obtain the first equivalence in (12). Therefore, we have that

$$\begin{aligned}
&\lambda\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts} \\
&\qquad\Downarrow \\
&\widehat{\boldsymbol{\Omega}}_{HB} = \boldsymbol{S}_{ct}\left[\lambda\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts}\right]\boldsymbol{S}_{ct}' \\
&\qquad\quad = \lambda\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,D}\boldsymbol{S}_{ct}' + (1-\lambda)\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts}\boldsymbol{S}_{ct}'.
\end{aligned}$$

The high-frequency time series representation (the second equivalence) can be derived in the following manner:

$$\begin{aligned}
\boldsymbol{\Omega} &= \boldsymbol{S}_{ct}\boldsymbol{\Omega}_{hf\text{-}bts}\boldsymbol{S}_{ct}' \\
&= (\boldsymbol{S}_{cs}\otimes\boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{S}_{te})' \\
&= (\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})'\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})' \\
&= (\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf}\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})'
\end{aligned}$$

where $\boldsymbol{\Omega}_{hf} = (\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})'$ and $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs}\otimes\boldsymbol{S}_{te} = (\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})$. We can apply the shrinkage estimator as

$$\begin{aligned}
&\lambda\widehat{\boldsymbol{\Omega}}_{hf,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{hf} \\
&\qquad\Downarrow \\
&\widehat{\boldsymbol{\Omega}}_{H} = (\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\left[\lambda\widehat{\boldsymbol{\Omega}}_{hf,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{hf}\right](\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})' \\
&\qquad\quad = \lambda(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\widehat{\boldsymbol{\Omega}}_{hf,D}(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})' + (1-\lambda)(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\widehat{\boldsymbol{\Omega}}_{hf}(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})'.
\end{aligned}$$

The bottom time series representation (the third equivalence) follows by

$$\begin{aligned}
\boldsymbol{\Omega} &= \boldsymbol{S}_{ct}\boldsymbol{\Omega}_{hf\text{-}bts}\boldsymbol{S}_{ct}' \\
&= (\boldsymbol{S}_{cs}\otimes\boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{S}_{te})' \\
&= (\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})'\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})' \\
&= (\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})\,\boldsymbol{\Omega}_{bts}\,(\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})',
\end{aligned}$$

where $\boldsymbol{\Omega}_{bts} = (\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})'$ and $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs}\otimes\boldsymbol{S}_{te} = (\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})\,(\boldsymbol{I}_n\otimes\boldsymbol{S}_{te})$. Finally we have that

$$\begin{aligned}
&\lambda\widehat{\boldsymbol{\Omega}}_{bts,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{bts} \\
&\qquad\Downarrow \\
&\widehat{\boldsymbol{\Omega}}_{B} = (\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})\left[\lambda\widehat{\boldsymbol{\Omega}}_{bts,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{bts}\right](\boldsymbol{S}_{cs}\otimes\boldsymbol{I}_{m+k^*})'
\end{aligned}$$

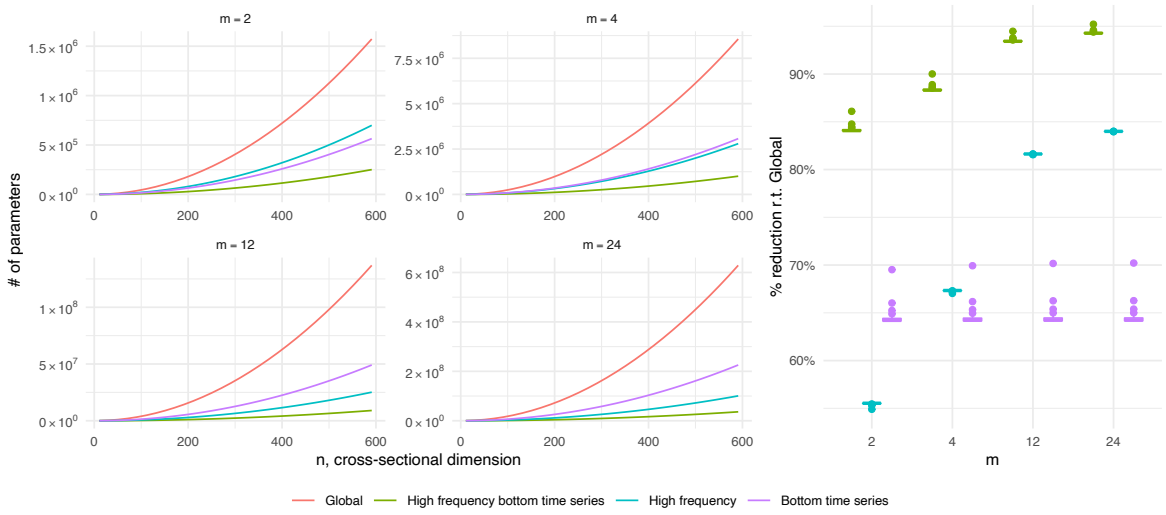

Figure 11: *The four graphs on the left represent the number of different parameters in the covariance matrix for the various approaches presented for different values of m and n (n_b, the number of bottom time series, is about 60% of the total). On the right, we have the boxplot of the percentage reduction in the number of parameters compared to the global approach.*

$$= \lambda \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right) \widehat{\boldsymbol{\Omega}}_{bts,D} \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right)' + (1-\lambda) \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right) \widehat{\boldsymbol{\Omega}}_{bts} \left(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*}\right)'.$$

In general, the covariance matrix of the reconciled forecasts is equal to $\boldsymbol{M}\widehat{\boldsymbol{\Omega}}\boldsymbol{M}'$ where $\boldsymbol{M} = \boldsymbol{S}_{ct}\boldsymbol{G}$ is the projection matrix. When using the HB approach, the covariance matrix of the reconciliation and the base forecasts will be identical. Indeed, it can be shown (see Panagiotelis et al. 2021 for more details) that if $\boldsymbol{M}$ is a projection matrix (7) then $\boldsymbol{M}\boldsymbol{S}_{ct} = \boldsymbol{S}_{ct}\boldsymbol{G}\boldsymbol{S}_{ct} = \boldsymbol{S}_{ct}$, and we obtain that

$$\begin{aligned}
\boldsymbol{M}\widehat{\boldsymbol{\Omega}}_{HB}\boldsymbol{M}' &= \boldsymbol{M}\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,HB}\boldsymbol{S}_{ct}'\boldsymbol{M}' \\
&= \boldsymbol{S}_{ct}\boldsymbol{G}\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,HB}\boldsymbol{S}_{ct}'\boldsymbol{G}'\boldsymbol{S}_{ct}' \\
&= \boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,HB}\boldsymbol{S}_{ct}' = \widehat{\boldsymbol{\Omega}}_{HB}.
\end{aligned}$$

Figure 11 shows the number of parameters for different values of m and n, with n_b fixed to approximately 60% of n. The right panel reports the boxplot of the percentage reductions in the number of parameters compared to the global approach.

C Cross-temporal covariance matrix for the Monte Carlo simulation in Section 5

We assume two AR(2) processes with correlated errors such that

$$y_{i,t} = \phi_{i,1}y_{i,t-1} + \phi_{i,2}y_{i,t-2} + \varepsilon_{i,t}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_2\left(\boldsymbol{0}_{(2\times 1)}, \boldsymbol{\Omega}_{cs}\right)$ and $i \in \{B, C\}$. Let $y_{i,T+h}$ be the true observation for the i^{th} series and $\widetilde{y}_{i,T+h}$ the corresponding forecasts such that

$$\begin{aligned} y_{i,T+1} &= \phi_{i,1}y_{i,T} + \phi_{i,2}y_{i,T-1} + \varepsilon_{i,T+1} \\ y_{i,T+2} &= \phi_{i,1}y_{i,T+1} + \phi_{i,2}y_{i,T} + \varepsilon_{i,T+2} \end{aligned} \quad \text{and} \quad \begin{aligned} \widetilde{y}_{i,T+1} &= \phi_{i,1}y_{i,T} + \phi_{i,2}y_{i,T-1} \\ \widetilde{y}_{i,T+2} &= \phi_{i,1}\widetilde{y}_{i,T+1} + \phi_{i,2}y_{i,T} \end{aligned},$$

then

$$
\begin{aligned}
y_{i,T+1} - \widetilde{y}_{i,T+1} &= \varepsilon_{i,T+1} \\
y_{i,T+2} - \widetilde{y}_{i,T+2} &= \varepsilon_{i,T+2} + \phi_{i,1}\varepsilon_{i,T+1}.
\end{aligned}
$$

Finally, we can compute each element of the high frequency bottom time series covariance matrix

$$
\begin{aligned}
Var\left(y_{i,T+1} - \widetilde{y}_{i,T+1}\right) &= \sigma_i^2, \quad \forall i \in \{B, C\} \\
Var\left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right) &= \sigma_i^2\left(1 + \phi_{i,1}^2\right), \quad \forall i \in \{B, C\} \\
Cov\left[\left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right), \left(y_{i,T+1} - \widetilde{y}_{i,T+1}\right)\right] &= Cov\left[\left(y_{i,T+1} - \widetilde{y}_{i,T+1}\right), \left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right)\right] \\
&= \phi_{i,1}\sigma_i^2, \quad \forall i \in \{B, C\} \\
Cov\left[\left(y_{i,T+1} - \widetilde{y}_{i,T+1}\right), \left(y_{j,T+1} - \widetilde{y}_{j,T+1}\right)\right] &= Cov\left[\left(y_{j,T+1} - \widetilde{y}_{j,T+1}\right), \left(y_{i,T+1} - \widetilde{y}_{i,T+1}\right)\right] \\
&= \sigma_{i,j}, \quad \forall i, j \in \{B, C\}, \quad i \neq j \\
Cov\left[\left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right), \left(y_{j,T+1} - \widetilde{y}_{j,T+1}\right)\right] &= Cov\left[\left(y_{j,T+1} - \widetilde{y}_{j,T+1}\right), \left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right)\right] \\
&= \phi_{i,1}\sigma_{i,j}, \quad \forall i, j \in \{B, C\}, \quad i \neq j \\
Cov\left[\left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right), \left(y_{j,T+2} - \widetilde{y}_{j,T+2}\right)\right] &= Cov\left[\left(y_{j,T+2} - \widetilde{y}_{j,T+2}\right), \left(y_{i,T+2} - \widetilde{y}_{i,T+2}\right)\right] \\
&= \sigma_{i,j}\left(1 + \phi_{i,1}\phi_{j,1}\right), \quad \forall i, j \in \{B, C\}, \quad i \neq j.
\end{aligned}
$$

In conclusion,

$$
\boldsymbol{\Omega}_{\textit{hf-bts}} = \begin{bmatrix}
\sigma_B^2 & & & \\
\phi_{B,1}\sigma_B^2 & \sigma_B^2\left(1 + \phi_{B,1}^2\right) & & \\
\sigma_{BC} & \phi_{B,1}\sigma_{BC} & \sigma_C^2 & \\
\phi_{C,1}\sigma_{BC} & \sigma_{BC}\left(1 + \phi_{B,1}\phi_{C,1}\right) & \phi_{C,1}\sigma_C^2 & \sigma_C^2\left(1 + \phi_{C,1}^2\right)
\end{bmatrix}
$$

and

$$
\boldsymbol{\Omega}_{ct} = \boldsymbol{S}_{ct}\boldsymbol{\Omega}_{\textit{hf-bts}}\boldsymbol{S}_{ct}'.
$$

D Geographic divisions of Australia

Table D.2: *Geographic divisions of Australia in States, Zones e Regions. Zones formed by a single region are highlighted in italics and not numbered.*

Series	Name	Label	Series	Name	Label
Total			*continues Regions*		
1	Australia	Total	49	Gippsland	BCB
States			50	Phillip Island	BCC
2	New South Wales (NSW)	A	51	Central Murray	BDA
3	Victoria (VIC)	B	52	Goulburn	BDB
4	Queensland (QLD)	C	53	High Country	BDC
5	South Australia (SA)	D	54	Melbourne East	BDD
6	Western Australia (WA)	E	55	Upper Yarra	BDE
7	Tasmania (TAS)	F	56	MurrayEast	BDF
8	Northern Territory (NT)	G	57	Mallee	BEA
Zones			58	Wimmera	BEB
9	Metro NSW	AA	59	Western Grampians	BEC
10	Nth Coast NSW	AB	60	Bendigo Loddon	BED
	Sth Coast NSW	*AC*	61	Macedon	BEE
11	Sth NSW	AD	62	Spa Country	BEF
12	Nth NSW	AE	63	Ballarat	BEG
	ACT	*AF*	64	Central Highlands	BEG
13	Metro VIC	BA	65	Gold Coast	CAA
	West Coast VIC	*BB*	66	Brisbane	CAB
14	East Coast VIC	BC	67	Sunshine Coast	CAC
15	Nth East VIC	BD	68	Central Queensland	CBA
16	Nth West VIC	BE	69	Bundaberg	CBB
17	Metro QLD	CA	70	Fraser Coast	CBC
18	Central Coast QLD	CB	71	Mackay	CBD
19	Nth Coast QLD	CC	72	Whitsundays	CCA
20	Inland QLD	CD	73	Northern	CCB
21	Metro SA	DA	74	Tropical North Queensland	CCC
22	Sth Coast SA	DB	75	Darling Downs	CDA
23	Inland SA	DC	76	Outback	CDB
24	West Coast SA	DD	77	Adelaide	DAA
25	West CoastWA	EA	78	Barossa	DAB
	Nth WA	*EB*	79	Adelaide Hills	DAC
	SthWA	*EC*	80	Limestone Coast	DBA
	Sth TAS	*FA*	81	Fleurieu Peninsula	DBB
26	Nth East TAS	FB	82	Kangaroo Island	DBC
27	Nth West TAS	FC	83	Murraylands	DCA
28	Nth Coast NT	GA	84	Riverland	DCB
29	Central NT	GB	85	Clare Valley	DCC
Regions			86	Flinders Range and Outback	DCD
30	Sydney	AAA	87	Eyre Peninsula	DDA
31	Central Coast	AAB	88	Yorke Peninsula	DDB
32	Hunter	ABA	89	Australia's Coral Coast	EAA
33	North Coast NSW	ABB	90	Experience Perth	EAB
34	South Coast	ACA	91	Australia's SouthWest	EAC
35	Snowy Mountains	ADA	92	Australia's North West	EBA
36	Capital Country	ADB	93	Australia's Golden Outback	ECA
37	The Murray	ADC	94	Hobart and the South	FAA
38	Riverina	ADD	95	East Coast	FBA
39	Central NSW	AEA	96	Launceston, Tamar and the North	FBB
40	New England North West	AEB	97	North West	FCA
41	Outback NSW	AEC	98	WildernessWest	FCB
42	Blue Mountains	AED	99	Darwin	GAA
43	Canberra	AFA	100	Kakadu Arnhem	GAB
44	Melbourne	BAA	101	Katherine Daly	GAC
45	Peninsula	BAB	102	Barkly	GBA
46	Geelong	BAC	103	Lasseter	GBB
47	Western	BBA	104	Alice Springs	GBC
48	Lakes	BCA	105	MacDonnell	GBD

Source: Wickramasuriya et al. (2019), Di Fonzo & Girolimetto (2022*b*)