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

23 February 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 non parametric bootstrap and a parametric Gaussian approach to draw samples from an incoherent cross-temporal distribution. The multi-step residuals are used for a better estimation of the covariance matrix, specifically in the time dimension where the usual one-step residuals fail. To address the 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 we introduce the idea of overlapping residuals. A simulation study is performed to investigate the theoretical and empirical proprieties of the different approaches. 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 these applications, the optimal cross-temporal reconciliation approaches significantly outperform the base forecasts according to the Continuous Ranked Probability Score and to the Energy Score. These findings demonstrate the effectiveness of the proposed techniques in improving the accuracy of probabilistic forecasting models. Overall, the paper expands and unifies the notation for cross-sectional, temporal and cross-temporal reconciliation and investigate the probabilistic cross-temporal framework in more detail.

Keywords: Probabilistic forecasting, Linearly constrained multiple time series, Cross-temporal framework, Forecast reconciliation

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 components, electricity demand and power generation, demand in supply chains with product categories, tourist arrivals and visitor nights separated by both 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 hierarchical structure by forecasting only one level. The bottom-up approach (Dunn et al. 1976) starts by summing up forecasts at the most disaggregated level to arrive at the desired aggregate level. 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 and forecast reconciliation have 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). The problem of obtaining consistent 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), and Zambon et al. (2022). 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 they 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: a non-parametric approach that bootstraps the base model residuals, and a parametric approach that assumes Gaussianity. 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 forecasting horizons when we deal with temporal levels, since one-step residuals are not suitable for this purpose. Alternative forms for constructing the covariance matrix and overlapping residuals are considered to solve high-dimensionality issues. 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 2022).

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 non-parametric bootstrap and a parametric Gaussian 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 Section 5, we explore cross-temporal residuals (overlapping and multi-step) looking at their advantages and limitations. A simulation study is performed in Section 6, to better understand the properties of the methodology. Two empirical applications using the Australian GDP and the Australian Tourism Demand datasets are presented in Sections 7 and 8, respectively. Finally, Section 9 presents conclusions and the research agenda on this and other related topics in the future.

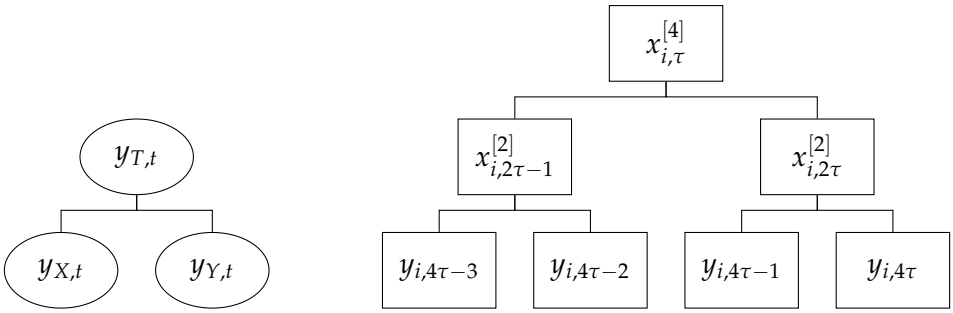
2 Notation and definitions

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.

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



(a) **(b)**

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

that $T = X + Y$. 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)

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

Looking at the hierarchical example in Figure 1(a), we have that

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

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 with $\mathcal{K} = \{k_p, k_{p-1}, \dots, k_2, k_1\}$ 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 aggregated 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 aggregation orders

$$\boldsymbol{x}_{i,\tau} = \begin{bmatrix} x^{[k_p]}_{i,\tau} & \boldsymbol{x}^{[k_{p-1}]\prime}_{i,\tau} & \dots & \boldsymbol{x}^{[1]\prime}_{i,\tau} \end{bmatrix}', \qquad \tau = 1, \dots, N.$$

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

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

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

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

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

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

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

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

where $\boldsymbol{x}^{[k]}_t = \left[\boldsymbol{u}^{[k]\prime}_t,\ \boldsymbol{b}^{[k]\prime}_t\right]'$ with $\boldsymbol{u}^{[k]}_t = \left[x^{[k]}_{1,t},\ldots,x^{[k]}_{n_a,t}\right]'$ is the n_a-vector of upper time series and $\boldsymbol{b}^{[k]}_t = \left[x^{[k]}_{(n_a+1),t},\ldots,x^{[k]}_{n,t}\right]'$ 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:

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

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

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

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 the zero-constrained representation through the $[(n_a m + nk^*) \times n(m + k^*)]$ zero constraints cross-temporal kernel 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

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

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 $\boldsymbol{b}^{[1]}_\tau = \text{vec}(\boldsymbol{B}^{[1]\prime}_\tau)$. We observe that,

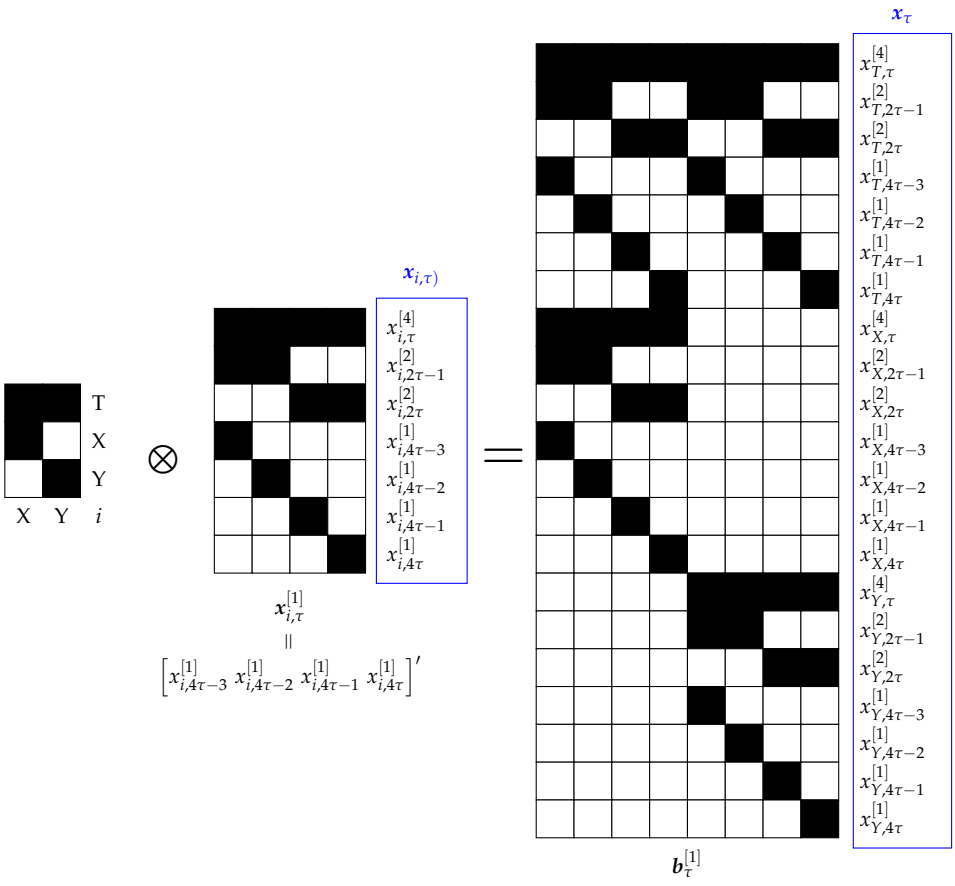


Figure 2: *Representation of the cross-temporal summing matrix $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te}$ defined in equation (6) for a system of 3 linearly constrained quarterly time series.*

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, as 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 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, that is:

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

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*)

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

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 $\widetilde{\boldsymbol{x}}_h = \text{vec}(\widetilde{\boldsymbol{X}}_h')$. It is worth noting that $\boldsymbol{C}_{ct}$ defined in equation (5) is a full rank matrix.

Wickramasuriya et al. (2019) showed that the minimum variance linear unbiased reconciled forecasts, satisfying

$$\min \ \text{tr}[\text{Var}(\widetilde{\boldsymbol{x}}_h - \boldsymbol{x})] \qquad \text{s.t. } \text{E}(\widetilde{\boldsymbol{x}}_h - \boldsymbol{x}) = 0,$$

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

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

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.

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: $\boldsymbol{\Omega}_{ct} = \text{diag}(\boldsymbol{S}_{ct}\mathbf{1}_{mn_b})$.

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

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

a block diagonal matrix where $\widehat{\boldsymbol{W}}^{[k]}_{shr}$ is the shrunk cross-sectional estimates of matrix proposed by Wickramasuriya et al. (2019) fixed the temporal aggregation level k.

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: $\boldsymbol{\Omega}_{ct} = \widehat{\boldsymbol{\Omega}}_{ct}$.

Table 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 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($struc$): $\boldsymbol{W} = \text{diag}(\boldsymbol{S}_{cs}\mathbf{1}_{nb})$	te($struc$): $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{S}_{te}\mathbf{1}_m)$
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 $\widehat{\boldsymbol{\Omega}}_D = \boldsymbol{I}_{k^*+m} \odot \widehat{\boldsymbol{\Omega}}$.

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 gives a visual representation of partly bottom-up as 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 the temporal bottom-up to each series' forecasts.

On the right (Figure 3b), we first compute the temporally reconciled forecasts for the most disaggregated cross-sectional level, and then apply the cross-sectional bottom-up approach. We denote this two-step reconciliation as, respectively, $\text{ct}(rec_{te}, bu_{cs})$, and $\text{ct}(rec_{cs}, bu_{te})$, where '$rec_{te}$' and '$rec_{cs}$' denote a generic forecast reconciliation approach in the temporal and in cross-sectional framework.

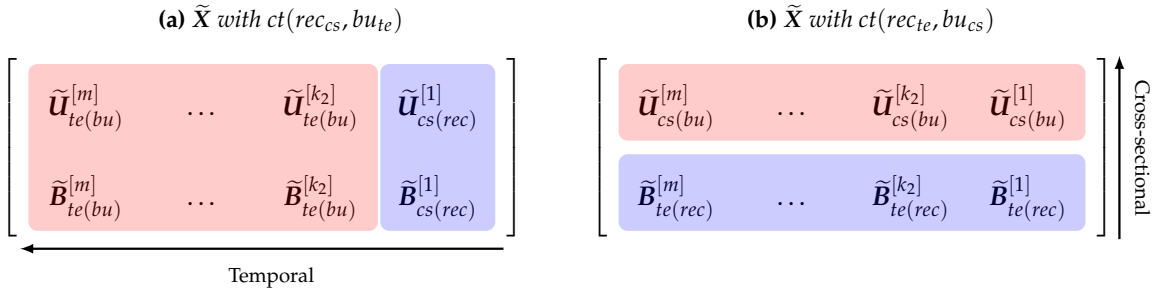


Figure 3: *A visual representation of partly bottom up starting from* (3a) *cross-sectionally reconciled forecasts for the temporal order 1 ($\widetilde{\boldsymbol{U}}^{[1]}$ and $\widetilde{\boldsymbol{B}}^{[1]}$) follow by a temporal bottom up, and* (3b) *temporally reconciled forecasts of the cross-sectional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

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 (2021) 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.

We want to extend these definitions to *cross-temporal coherent probabilistic forecasts* and *cross-temporal probabilistic forecast reconciliation*. Let $(\mathbb{R}^{n_b m}, \mathcal{F}_{\mathbb{R}^{n_b m}}, \mu)$ be a probability space for the bottom time series $\boldsymbol{b}^{[1]}_{\tau}$, 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 $(\mathbb{R}^{n_b m}, \mathcal{F}_{\mathbb{R}^{n_b m}}, \mu)$, we define the coherent probability space as the triple $(\mathfrak{s}, \mathcal{F}_{\mathfrak{s}}, \breve{\mu})$ satisfying the following property:

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

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

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

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

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 of $s \circ g$, as for the cross-temporal point reconciliation (8).

In a similar way, the definitions and theorems presented in Corani et al. (2021) can be applied to the cross-temporal case as well.

3.1 Non-parametric framework: bootstrap reconciliation

Analytical expressions for the base and reconciled forecast distributions are sometimes challenging to express, and they sometimes use unrealistic parametric assumptions.

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 $\widehat{\nu}$. 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 $\widetilde{\nu}$ as 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.

An important issue is how to generate samples from the base forecast distributions that preserve cross-temporal relationships. We suggest the cross-temporal joint (block) bootstrap (**ctjb**) to solve this issue. The 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 4 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 historical series.

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

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

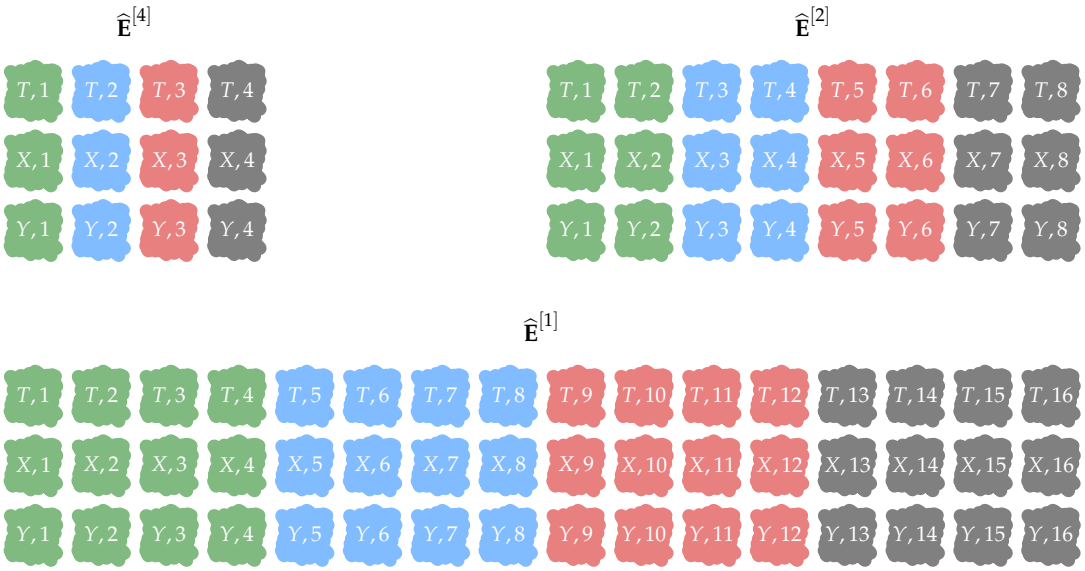


Figure 4: *Example of residual matrices for figure 3 with 4 years of data ($N = 4$): the green color corresponds to the first year, the blue to the second year, the red to the third year and the black to the fourth year.*

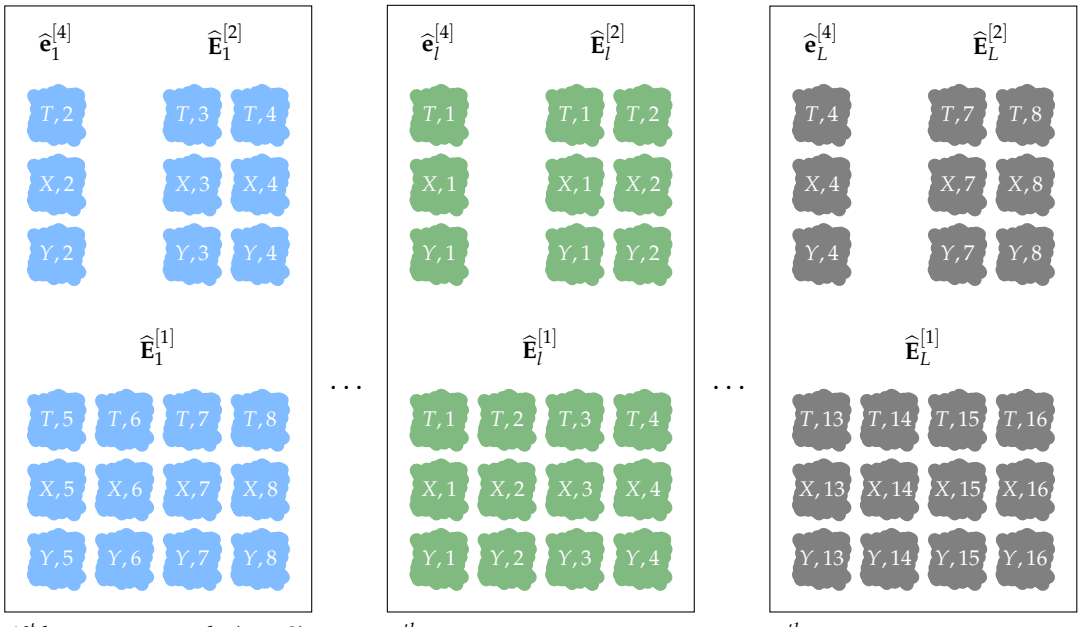


Figure 5: *Example of bootstrapped residuals for the cross-temporal hierarchy of Figure 1 and using the residuals in Figure 4.*

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

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

where $\widehat{\boldsymbol{E}}_\tau = \begin{bmatrix} \widehat{\boldsymbol{e}}_\tau^{[m]} & \widehat{\boldsymbol{E}}_\tau^{[k_{p-1}]} & \dots & \widehat{\boldsymbol{E}}_\tau^{[1]} \end{bmatrix}$ is a $[n \times (k^* + m)]$ matrix where $\widehat{\boldsymbol{e}}_\tau^{[m]}$ is the τ^{th} column of $\widehat{\boldsymbol{E}}^{[m]}$ and $\widehat{\boldsymbol{E}}_\tau^{[k]}$ is the $(n \times M_k)$ matrix formed by columns $M_k(\tau-1)+1, \dots, M_k\tau$ of $\widehat{\boldsymbol{E}}^{[k]}$. Figure 5 shows the component of $\widehat{\boldsymbol{E}}_\tau = \begin{bmatrix} \widehat{\boldsymbol{e}}_\tau^{[4]} & \widehat{\boldsymbol{E}}_\tau^{[2]} & \widehat{\boldsymbol{E}}_\tau^{[1]} \end{bmatrix}$ 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 cross-sectional series. This can be especially useful when dealing with large datasets or when trying to speed up the analysis process.

3.2 Parametric framework: Gaussian reconciliation

It is possible to obtain a reconciled probabilistic forecast analytically for some parametric distributions, such as the multivariate normal (Panagiotelis et al. 2023, Wickramasuriya 2021, Corani et al. 2021, Eckert et al. 2021). 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 matrix notation in Section 2, we may extend this results to the cross-temporal case. To obtain a reconciled forecast using the multivariate normal distribution for given $H = 1$, 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

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

where $\boldsymbol{M}$ is the projection matrix defined in equation (7). Note that if we assume that $\boldsymbol{\Sigma} = \boldsymbol{\Omega}_{ct}$, then the covariance matrix in equation (10) simplifies to $\widetilde{\boldsymbol{\Omega}} = \boldsymbol{M}\boldsymbol{\Omega}_{ct}$. In the cross-temporal case, accurately 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 incorrect estimate of the covariance matrix. These challenges will be explored in more depth in the following sections.

Focusing on the computational aspect, we can take several steps to reduce the time required to obtain simulations from the reconciled forecast distribution. For example, it is not necessary to simulate from a normal distribution with a defined covariance matrix for the entire structure when dealing with a genuine hierarchical 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 formula (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 6.

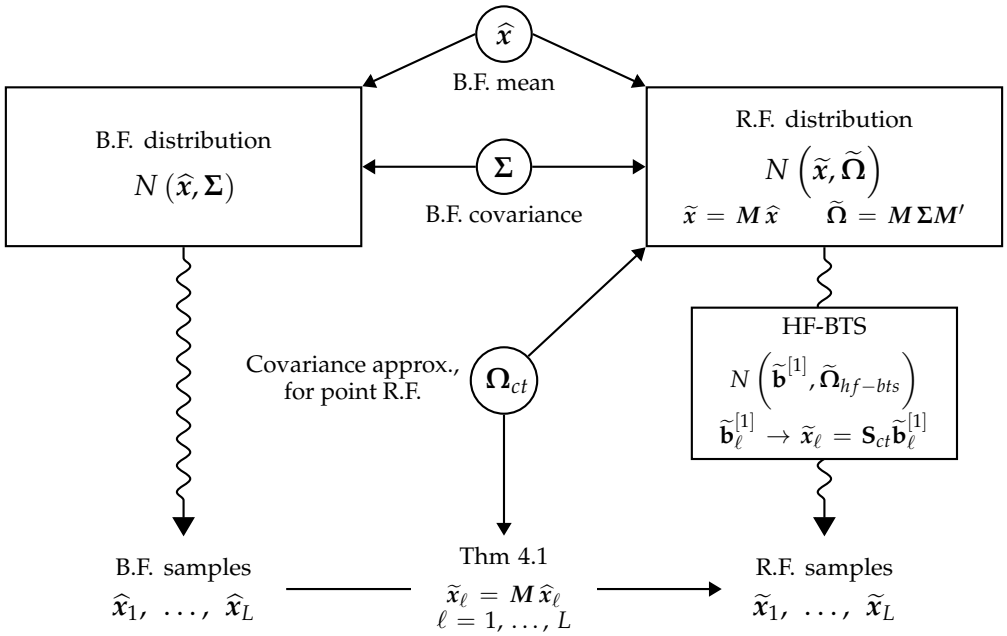


Figure 6: *Summary diagram of reconciliation in the Gaussian framework, as described in Section 3.2. The acronyms R.F and B.F. stand for Reconciled Forecasts and Base Forecasts, respectively. HF-BTS stands for High Frequency Bottom Time Series.*

4 Shrinkage techniques for cross-temporal covariance matrix estimation

In this section, our focus will be exclusively on the cross-temporal framework, and so we will drop the *ct* subscript in order to simplify the notation.

As the covariance matrix $\boldsymbol{\Omega}$ is unknown in practice, a natural estimate is the empirical sample covariance matrix of the base forecasts $\widehat{\boldsymbol{\Omega}}$. In 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 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. Schäfer & Strimmer (2005) proposed an unbiased estimator for the shrinkage intensity parameter, given by

$$\widehat{\lambda} = \frac{\sum_{i \neq j} \widehat{\text{Var}}(\widehat{\sigma}_{ij})}{\sum_{i \neq j} \widehat{\sigma}_{ij}^2},$$

where $\widehat{\sigma}_{ij}$ is the ij^{th} element of $\widehat{\mathbf{\Omega}}$. In finite samples, $\widehat{\lambda}$ may exceed unity or even become negative; to avoid over-shrinkage or negative shrinkage, the condition

$$\widehat{\lambda}^* = \max\left[0, \min(1, \widehat{\lambda})\right]$$

is imposed to ensure that the estimated λ stays within the range $[0,1]$. The two matrices obtained with $\lambda = 0$ and $\lambda = 1$ respectively are shown in the first column of Figure 7. 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 Section 2.1.

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. We consider three different approaches at this purpose.

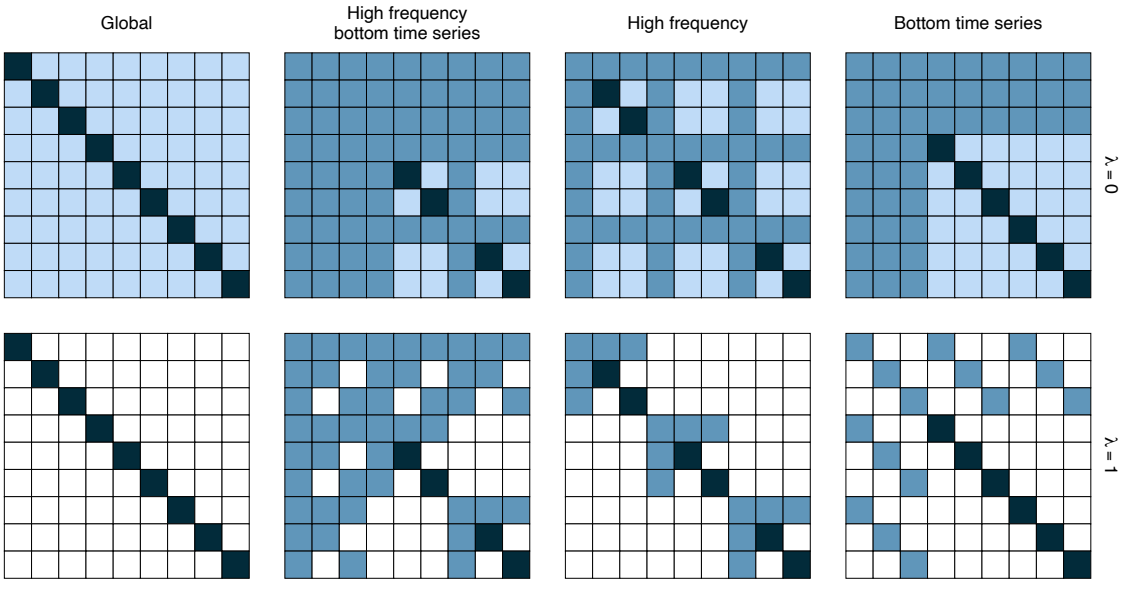


Figure 7: *Representation of four types of covariance matrices that can be obtained from the cross-temporal hierarchical structure (3 time series and $m = 2$) for two different values of $\lambda \in \{0,1\}$, the shrinkage parameter. The cells that are not modified by shrinkage are colored black, those 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.*

The first idea is to use both cross-sectional and temporal structure simultaneously. In general we know that the true covariance matrix can be written as

$$\boldsymbol{\Omega} = \boldsymbol{S}_{ct}\boldsymbol{\Omega}_{hf\text{-}bts}\boldsymbol{S}'_{ct} \tag{12}$$

where $\boldsymbol{\Omega}_{hf\text{-}bts}$ is the covariance matrix for the bottom time series at aggregation level $k = 1$ (highest frequency bottom time series). Therefore, we can apply the idea of "Stein-type shrinkage" to $\boldsymbol{\Omega}_{hf\text{-}bts}$ by using the empirical covariance matrix of the high frequency bottom base forecasts $\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts}$, such that

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

and

$$\begin{aligned}\widehat{\boldsymbol{\Omega}}_{HB} &= \boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,HB}\boldsymbol{S}'_{ct} \\ &= \lambda\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,D}\boldsymbol{S}'_{ct} + (1-\lambda)\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts}\boldsymbol{S}'_{ct},\end{aligned}$$

where $\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,D} = \boldsymbol{I}_{n_b m} \odot \widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,HB}$ is a diagonal matrix, λ is the shrinkage parameter that can be estimate using Schäfer & Strimmer (2005) and $\widehat{\boldsymbol{\Omega}}_{HB}$ is the cross-temporal covariance matrix for the complete system, with HB denoting High frequency Bottom time series shrinkage. When using this approach, the covariance matrix of the reconciliation and the base forecasts will be identical. In general, the covariance matrix of the reconciled forecasts is equal to $\boldsymbol{M}\widehat{\boldsymbol{\Omega}}_{HB}\boldsymbol{M}'$ where $\boldsymbol{M} = \boldsymbol{S}_{ct}\boldsymbol{G}$ is the projection matrix. 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

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

However, when working with general linearly constrained multiple time series (where a natural hierarchy representation is not available), we may prefer not to impose the cross-sectional structure. Starting from (12), we know that

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

where $\boldsymbol{\Omega}_{hf} = (\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})'$ is the covariance matrix related to all the high frequency time series 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 estimate the matrix $\boldsymbol{\Omega}_{hf}$ using the high frequency base forecasts $\widehat{\boldsymbol{\Omega}}_{hf}$. Therefore,

$$\widehat{\boldsymbol{\Omega}}_{hf,H} = \lambda\widehat{\boldsymbol{\Omega}}_{hf,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{hf}$$

and

$$\widehat{\boldsymbol{\Omega}}_{H} = (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})\widehat{\boldsymbol{\Omega}}_{hf,H}(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})'$$

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

where $\widehat{\boldsymbol{\Omega}}_{hf,D} = \boldsymbol{I}_{nm} \odot \widehat{\boldsymbol{\Omega}}_{hf,HB}$ is a diagonal matrix, and $\widehat{\boldsymbol{\Omega}}_H$ is the cross-temporal covariance matrix for the complete system, with H denoting High frequency shrinkage.

Finally, we can impose only the cross-sectional structure. We know that

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

where $\boldsymbol{\Omega}_{bts} = (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf\text{-}bts}\,(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})'$ is the covariance matrix related to bottom time series at any temporal aggregation, and $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te} = (\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})\,(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})$. We can estimate the matrix $\boldsymbol{\Omega}_{bts}$ using the bottom time series base forecasts $\widehat{\boldsymbol{\Omega}}_{bts}$:

$$\widehat{\boldsymbol{\Omega}}_{bts,B} = \lambda\widehat{\boldsymbol{\Omega}}_{bts,D} + (1-\lambda)\widehat{\boldsymbol{\Omega}}_{bts}$$

and

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

where $\widehat{\boldsymbol{\Omega}}_{bts,D} = \boldsymbol{I}_{n_b(m+k^*)} \odot \widehat{\boldsymbol{\Omega}}_{bts,B}$ is a diagonal matrix, and $\widehat{\boldsymbol{\Omega}}_B$ is the cross-temporal covariance matrix for the complete system, with B denoting Bottom time series shrinkage.

These four approaches differ not only in their target matrices (see Figure 7), but also in the number of parameters to be estimated through the residuals of the base forecasts.

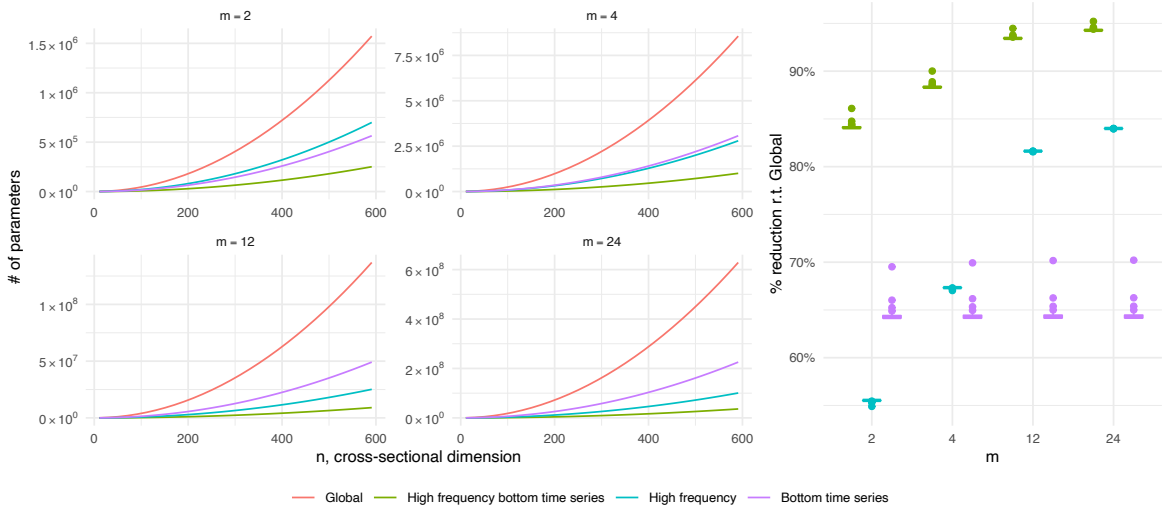


Figure 8: *The four graphs on the left represent the number of different parameters in the covariance matrix for the various approaches presented for different values of m and n (with n_b, the number of bottom time series, is about 60% of the total). On the right, we have the boxplot of the percentage reduction in the number of parameters compared to the global approach.*

Table 2 shows the total number of different parameters in the covariance matrix, while Figure 8 shows these values for different values of m and n, with n_b fixed to approximately 60% of n. As we can see, a global approach (G) usually involves a considerable number of parameters compared to other procedures. The right panel of Figure 8 reports the boxplot of the percentage reductions in the number of parameters compared to the global approach, that is:

$$\% \text{ reduction} = 100(1 - r_i/r_G) \quad \text{with } i \in \{HB, H, B\},$$

we observe that using only the high frequency bottom time series (HB) leads to a decrease of around 90%. The H and B approaches, on the other hand, are a compromise between the previous two. Although there is no clear hierarchy between the two, as m and n increase, using the high frequency time series requires the estimation of a smaller number of parameters. In Table 3 we report the number of different parameters for the Monte Carlo simulation

Shrinkage	**ID**	**# of different parameters**	
Global	G	$r = \frac{n(k^*+m)[n(k^*+m)-1]}{2}$	
High frequency bottom time series	HB	$r_{HB} = \frac{n_b m[n_b m-1]}{2}$	$r_{HB} < r$
High frequency	H	$r_H = \frac{nm[nm-1]}{2}$	$r_{HB} < r_H < r$
Bottom time series	B	$r_B = \frac{n_b(k^*+m)[n_b(k^*+m)-1]}{2}$	$r_{HB} < r_B < r$

Table 2: *Number of different parameters that need to be estimated for the various approaches. The Shrinkage column indicates the type of shrinking applied, and the ID column shows the abbreviations used.*

Shrinkage	**ID**	**AR(2)**	**AusGDP**	**VN525**
Global	G	36	221445	108052350
High frequency bottom time series	HB	6	30876	6655776
High frequency	H	15	72390	19848150
Bottom time series	B	15	94395	36231328

Table 3: *Number of different parameters that need to be estimated for the Monte Carlo simulation (**AR(2)**, see Section 6), the Australian GDP (**AusGDP**, see Section 7) dataset and the Australian Tourism Demand (**VN525**, see Section 8): 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).*

(**AR2**), the Australian Tourism Demand (**VN525**) and the Australian GDP (**AusGDP**) datasets considered in the next sections.

In our simulations and empirical analysis, 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 sample of the linearly constrained time series when we assume normality in our sample. Additionally, we will use the shrinkage versions as approximations for reconciliation. This will allow us to have a more comprehensive understanding of the data and the relationships between the series.

5 Residual analysis

5.1 Multi-step residuals

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

$$e^{[k]}_{i,h,t} = x^{[k]}_{i,t+h} - \widehat{x}^{[k]}_{i,t+h|t},$$

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

Following Di Fonzo & Girolimetto (2023*a*), we use a matrix organization of residuals that is similar to the one seen for 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, with $N_k = Nm/k$ 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

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

5.2 Overlapping residuals

Another issue that arises in the case of cross-temporal reconciliation is the low number of residuals available, especially for the higher levels 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

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

with $s = 0, \ldots, (k-1)$. 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:

$$\boldsymbol{x}^{[2],0} = \left[x_1^{[2],0},\ x_2^{[2],0},\ x_3^{[2],0} \right]' = \left[y_1 + y_2,\ y_3 + y_4,\ y_5 + y_6 \right]'$$

and

$$\boldsymbol{x}^{[2],1} = \left[x_1^{[2],1},\ x_2^{[2],1} \right]' = \left[y_2 + y_3,\ y_4 + y_5 \right]'.$$

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 $s = 1, \ldots, k-1$, 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 equation (14).

The resulting residuals can be used to estimate the covariance matrix in cross-temporal reconciliation. Using overlapping residuals allows for an increase in 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.

6 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 structure, which can provide valuable insights into the underlying dynamics of the time series data.

Consider a 2-level hierarchical structure with three time series in total: one upper series and two bottom series, with semi-annual data ($\mathcal{K} = \{2, 1\}$). The cross-sectional aggregation matrix is $\boldsymbol{A}_{cs} = \begin{bmatrix} 1 & 1 \end{bmatrix}$. This means that the upper series, A, is a combination of the two bottom series, B and C, such that $A = B + C$. To ensure a realistic bivariate series, the data

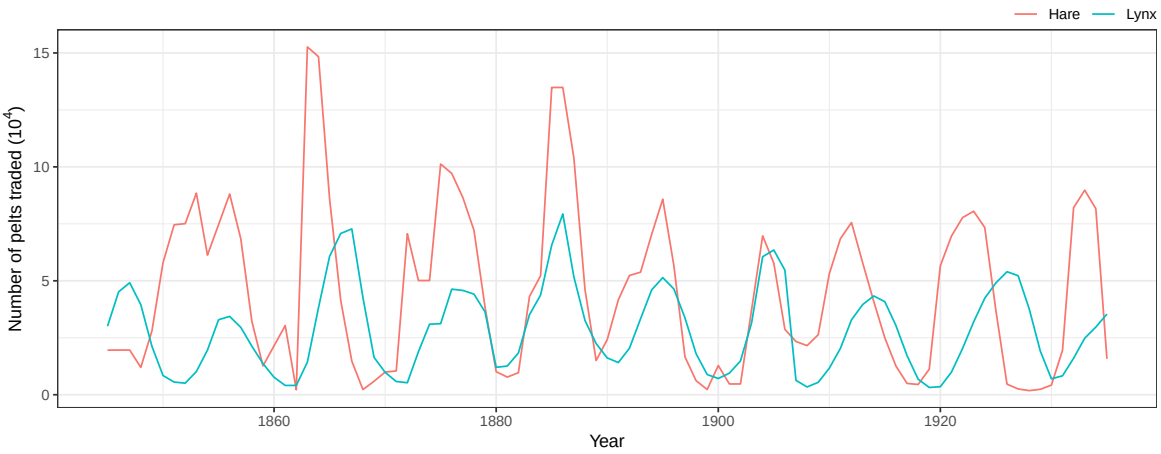


Figure 9: *The Hudson Bay Company kept records of their trading (from all the company's operating regions) in Snowshoe Hare and Canadian Lynx furs from 1845 to 1935. The "pelt" dataset is available in R with the* `tsibbledata` *package (O'Hara-Wild et al. 2022).*

for the bottom series B and C are generated by AR(2) processes, where the AR parameters are estimated from the "Lynx" and "Hare" time series shown in Figure 9 and contained in the `pelt` dataset in the `tsibbledata` package for R (O'Hara-Wild et al. 2022). The AR parameters for B and C are $\phi_B = [1.34, -0.74]'$ and $\phi_C = [0.95, -0.42]'$, respectively. The errors driving the AR processes are 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 given by

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

The bottom series are then cross-temporally aggregated to obtain the series at any temporal aggregation.

For the forecast experiment, base forecasts using AR models will be used. The order of the AR models is not fixed at the true order, but rather is determined using the algorithm proposed by Hyndman & Khandakar (2008) in the R package `forecast` (Hyndman et al. 2023), thus allowing for the possibility of mis-specification in the models. The training window length is 500 years, consisting of 1000 observations. The experiment will be replicated 500 times, with a forecast horizon of 1 year.

Since the models used for the data generating process (DPG) for the bottom series B and C at the most aggregated temporal level are two AR(2) models, we have calculated the true covariance matrix for one-step ahead forecasts at the annual level. Therefore, we have that $\mathbf{\Omega}_{ct} = \mathbf{S}_{ct}\mathbf{\Omega}_{hf\text{-}bts}\mathbf{S}'_{ct}$ where

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

The detailed calculations can be found in Appendix A. In Figure 10 we have represented both the covariance matrix and the correlation matrix for fixed parameters.

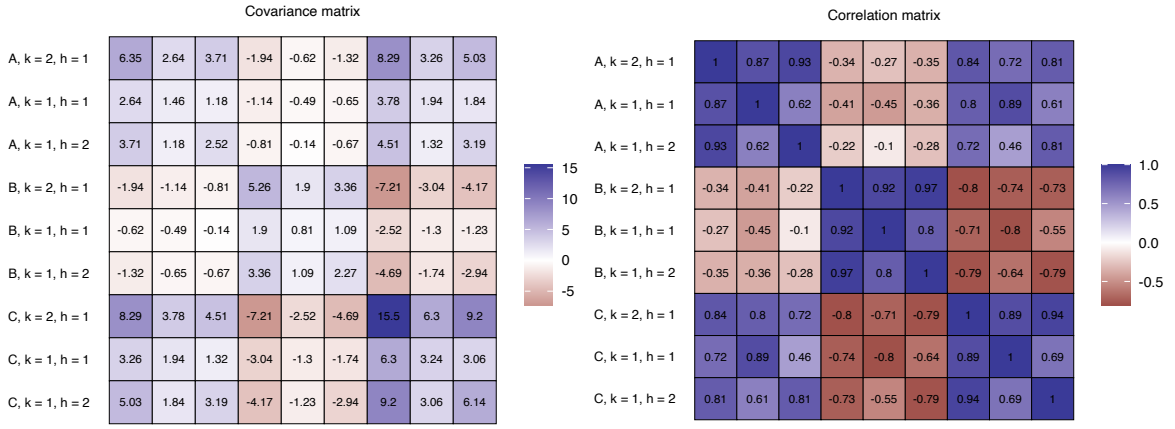


Figure 10: *True cross-temporal covariance (left) and correlation (right) matrix of forecasts reconciled with* $\sigma_B = 0.9$, $\sigma_C = 1.8$, $\phi_B = [1.34, -0.74]'$, $\phi_C = [0.95, -0.42]'$ *and* $\rho = -0.8$.

6.1 Base and reconciled samples

To construct cross-temporal samples of the base forecasts, we use the bootstrap and Gaussian approaches discussed in Sections 3.1 and 3.2, respectively. For the non-parametric approach, we use regular one-step residuals, while for the parametric approach we use multi-step residuals with the different covariance matrix structures analyzed in Section 4: G for the global covariance matrix, H for the High frequency covariance matrix, B for the bottom time series covariance matrix and HB for the high frequency bottom time series covariance matrix. We do not use overlapping residuals in our analysis because the large number of available observations made it unnecessary

We have several reconciliation options: cross-temporal bottom-up ($ct(bu)$); cross-sectional reconciliation using the *shr* approach (see Table 1) and then temporally aggregated (ct(shr_{cs}, bu_{te}), Section 2.2); temporal reconciliation using the *wlsv* approach (see Table 1) and then cross-sectionally aggregated (ct$(wlsv_{te}, bu_{cs})$, Section 2.2); optimal cross-temporal reconciliation (Di Fonzo & Girolimetto 2023*a*) with one-step residuals (**oct**) using the *wlsv*, *bdshr* approaches presented in Section 2.1. In addition, we propose the optimal cross-temporal reconciliation with multi-step residuals (**oct**$_h$) for *shr*, *hshr*, *bshr* and *hbshr* approaches that use, respectively, the Global, High frequency, Bottom time series, High frequency Bottom time series shrinkage presented in Section 4. All procedures are available in the R package `FoReco` (Girolimetto & Di Fonzo 2022).

6.2 A look at the covariance matrix

Figure 11 compares the estimated covariance and correlation matrices for base forecasts using different approaches. The first column represents the covariance matrix and the second column represents the correlation matrix. The first row uses a non-parametric bootstrap approach, while the second use a parametric Gaussian approach with multi-step residuals. The true covariance and correlation matrices are shown in Figure 10 for comparison. As

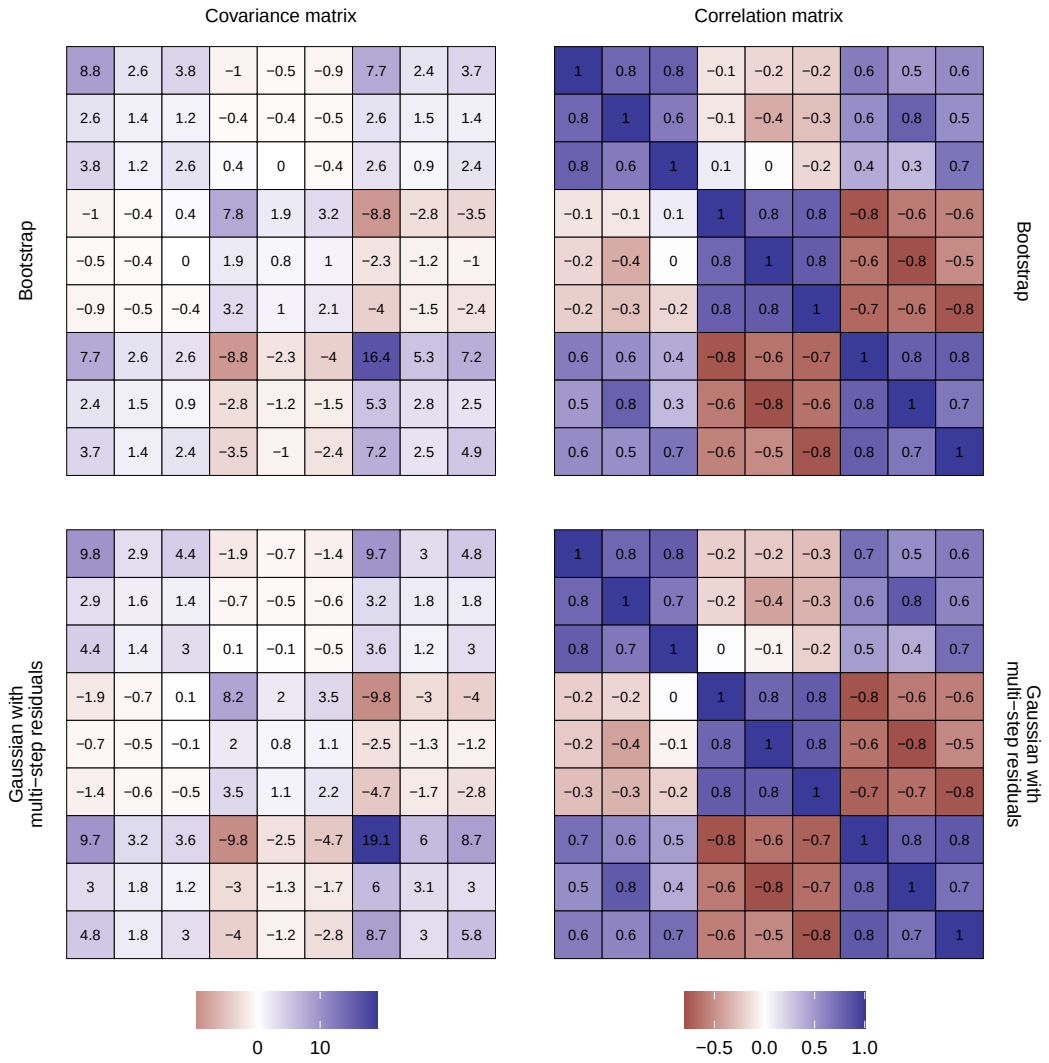


Figure 11: *Comparison of estimated covariance and correlation matrices (first simulation) for base forecasts using non-parametric bootstrap (first row) and parametric Gaussian (with multi-step residuals, second row) approaches. The true covariance and correlation matrices are shown in Figure 10.*

discuss in Section 5.1, if we use the one-step residuals in the gaussian framework we estimate a biased covariance matrix such that the correlations between the one-step and two-step ahead forecasts are almost always zero. However, we solve this issue using the multi-step residuals. On the other hand, the bootstrap approach (with one-step residuals) is not affected by the null correlation.

To compare the true covariance matrix $\mathbf{\Omega}_{ct}$ with the estimated covariance matrix $\mathbf{\Omega}$, we use the Frobenius norm that is a metric that quantifies 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 $\boldsymbol{D} = \mathbf{\Omega} - \mathbf{\Omega}_{ct}$. The true covariance matrix, shown in Figure 10, was compared to the estimated covariance matrix obtained using various reconciliation methods and techniques for base forecasts' samples. By comparing these two matrices using the Frobenius norm, we were able to determine which reconciliation method and simulation technique produce an

Reconciliation approach	Base forecasts' sample approach				
	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 4: *Frobenius norm between the true (in Figure 10) and the estimated covariance matrix for different reconciliation approaches and different techniques for simulating the base forecasts. In bold, it is reported the lowest value for each column, in blue the minimum. The notation used to refer to the reconciliation and base forecast samples is explained in more detail in Section 6.1.*

accurate estimate of the covariance matrix. Other types of matrix norms were also used with similar results.

From Table 4, it appears that the reconciled covariance matrices are always closer to the true matrix than the base forecast matrix when using both bootstrap and gaussian approach. Overall, there are no difference in the findings when using either one-step or multi-step residuals in cross-temporal 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 done. It is worth noting that if we simulate the base forecasts using the HB covariance matrix, then reconciliation does not modify it (see Section 4). In conclusion, our results suggest that using multi-step residuals or bootstrap techniques may help find an estimate of the covariance matrix close to its true counterpart, and that reconciliation can further improve this estimate.

6.3 Accuracy scores

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

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

where $\widehat{P}_i(\omega) = \frac{1}{L}\sum_{l=1}^{L}\mathbf{1}\left(x_{i,l} \leq \omega\right)$, $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_L \in \mathbb{R}^n$ is a collection of L random draws taken from the predictive distribution and $\boldsymbol{z} \in \mathbb{R}^n$ is the observation vector. This scoring rule is implemented in the R package `scoringRules` (Jordan et al. 2019). CRPS 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

(Wickramasuriya 2021, Panagiotelis et al. 2023) ,

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

In Tables 5 and 6 are reported the results using these two indices. In particular, we consider the geometric mean of the skill scores of the CRPS indices (Fleming & Wallace 1986), and the standard skill score for index ES:

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

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

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

and

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

A limitation of this simulation experiment is that we are using a high number of residuals, which may result in a potential loss of benefit when using forms of covariance matrix parameterization 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 $\mathrm{ct}(bu)$ (cross-temporal bottom-up) approach is a reconciliation method that starts by base forecasts at the lowest level of aggregation (bottom base forecasts for $k = 1$), and then aggregates the reconciled forecasts up to the highest level of aggregation. The good performance of the $\mathrm{ct}(bu)$ approach can be explained by a good quality of 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 $\mathrm{ct}(shr_{cs}, bu_{te})$, 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, there does not seem to be any advantage in using multi-step residuals to calculate the covariance matrix in the reconciliation process, as the value of the index is not significantly different when using one-step or multi-step residuals in this case.

In conclusion, we explored the use of simulation and reconciliation techniques to improve the accuracy of forecasts. We found that simulating base forecasts from multi-step residuals

Reconciliation approach	Base forecasts' sample approach									
	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 5: *CRPS skill score defined in* (17) *and* (18). *The smaller this value, the more accurate the forecast. Approaches that performed worse than the benchmark model (Bootstrap base forecasts) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The notation used to refer to the reconciliation and base forecast samples is explained in Section 6.1.*

estimates a covariance matrix closed to the true one and adjusts the null correlation when simulating from one-step residuals. Additionally, we observed that reconciliation could be used to further improve the accuracy of these estimates. We obtained good base forecasts for $k = 1$, which give us good performance for bottom-up techniques. Optimal cross-temporal reconciliation techniques, such as oct($wlsv$) and oct($bdshr$), which assume either zero temporal/cross-sectional correlations or a cross-sectional correlation structure equal for each level of temporal aggregation, respectively, were able to achieve good results in terms of both CRPS and ES.

7 Australian GDP dataset

The Australian Quarterly National Accounts dataset (AusGDP) has been widely used in the literature for cross-sectional and cross-temporal reconciliation. 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

Reconciliation approach	Base forecasts' sample approach									
	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 6: *ES skill score defined in equation* (17) *and* (19). *The smaller this value, the more accurate the forecast. Approaches that performed worse than the benchmark model (Bootstrap base forecasts) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The notation used to refer to the reconciliation and base forecast samples is explained in Section 6.1.*

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*) to point reconciliation in a cross-temporal framework.

Building on the results of this analysis, probabilistic cross-temporal forecast reconciliation is now applied within the same forecasting experiment designed by Di Fonzo & Girolimetto (2023*a*), which is the cross-temporal extension of the cross-sectional forecasting experiment in Athanasopoulos et al. (2020). Univariate ARIMA models were used 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). The forecasting experiment used a recursive training sample with an expanding window: the first training sample spanned from 1984:Q4 to 1994:Q3, and the last ended on 2017:Q1, for a total of 91 forecast origins. In addition, one-step-ahead and two-step-ahead forecasts were also

computed for time series obtained by aggregating two successive quarters and one-step-ahead forecasts were computed for time series obtained by aggregating four successive quarters.

7.1 Base and reconciled samples

To construct the base forecast samples in the Gaussian case, we used the *Global* (G) and *High frequency* (H) sample covariance matrices presented in Section 4 (the results with shrinkage version are available in the online appendix), since it is not possible to identify a unique representation (Di Fonzo & Girolimetto 2022*d*) for the other two structures *High frequency Bottom time series* (HB) and *Bottom time series* (B). We will 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 3.1. Finally, to reconcile the resulting (1000) base forecasts samples, we have applied the following techniques using the R package `FoReco` (Girolimetto & Di Fonzo 2022):

- $\mathbf{ct}(\cdot, bu_{te})$ partly bottom-up (Section 2.2) starting from cross-sectional reconciled forecasts using the *shr* and *wls* approaches (Table 1);
- $\mathbf{oct}_{o}(\cdot)$ optimal cross-temporal reconciliation with overlapping residuals for the *wlsv* (Section 2.1) and *bdshr* (Section 2.1);
- $\mathbf{oct}_{oh}(\cdot)$ optimal cross-temporal reconciliation with overlapping and multi-step residuals for the *shr* (*Global shrinkage* in Section 4) and *hshr* (*High frequency shrinkage* in Section 4) approaches.

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).

To accurately evaluate the quality of our forecasts, we used the Continuous Ranked Probability Score (CRPS) and the Energy Score (ES). These measures are detailed in formulas (15) and (16), respectively, along with their corresponding skill scores (17), (18), and (19). In order to provide a comprehensive understanding of our evaluation results, we will present and analyze these accuracy indices at multiple temporal levels. This will allow us to examine the performance of our forecasting techniques at different durations, giving us a more complete picture of their effectiveness. Additionally, we utilized 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.

7.2 Results

In Tables 7 and 8, the skill scores for CRPS (Continuous Ranked Probability Score) and ES (Energy Score) are presented, respectively. As a benchmark for calculating skill scores, we used the base forecasts calculated using the bootstrap method. At the base forecasts level, we found that using a parametric approach with the normal distribution performs better than the non-parametric approach. This is likely due to the nature of the data and the limited number

Reconciliation approach	Base forecasts' sample approach									
	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 7: *CRPS skill score defined in equation* (17) *and* (18) *for the Australian Quarterly National Accounts dataset (AusGDP). The smaller this value, the more accurate the forecast. Approaches that performed worse than the benchmark model (Bootstrap base forecasts) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The notation used to refer to the reconciliation and base forecast samples is explained in Section 7.1.*

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 techniques. Among all the procedures, $ct(wls_{cs}, bu_{te})$ and $oct_o(wlsv)$ showed the greatest relative gains. In contrast, $oct_{oh}(shr)$ and $oct_{oh}(hshr)$ techniques seemed to have more difficulty to improve. Furthermore, the greatest improvements were observed for higher aggregation times.

Figure 12 shows the base (first row) and reconciled (second and third rows) forecasts for Australian GDP at the second iteration of our forecasting experiment for the bootstrap and Gaussian approaches in the first and second columns respectively. The green shading represents the 80% confidence interval, while the black line with the triangle represents the observed true values. The red line and dot represent the median value, and the blue line and dot represent the mean value. In the nonparametric case, the median and mean are not always overlapped and centered within the confidence interval (asymmetrical), as is the case in the Gaussian distribution. This is because the nonparametric case does not assume a specific underlying distribution for the data.

Figure 13 compare the MCB using the CRPS for the non-parametric bootstrap approach and the Gaussian approach using overlapping and multi-step residuals, with the covariance matrix calculated from the matrix of high frequency time series (H). Also in this case, $ct(wls_{cs}, bu_{te})$ and $oct_o(wlsv)$ were found to be significantly better than base forecasts at any level of aggregation.

Reconciliation approach	Base forecasts' sample approach									
	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 8: *ES skill score defined in equation* (17) *and* (19) *for the Australian Quarterly National Accounts dataset (AusGDP). The smaller this value, the more accurate the forecast. Approaches that performed worse than the benchmark model (Bootstrap base forecasts) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The notation used to refer to the reconciliation and base forecast samples is explained in Section 7.1.*

Overall, we found 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 base forecasts take into account the multi-step structure.

8 Australian Tourism Demand dataset

The Australian Tourism Demand dataset (Wickramasuriya et al. 2019, 2020) measures the number of tourists arriving and spending nights in tourist facilities. It includes 228 monthly observations of Visitor Nights (VN) from January 1998 to December 2016, and has a cross-sectional clustered structure based on a geographic hierarchy and a disaggregation by purpose of travel. The geographic hierarchy is divided into seven states, 27 zones of states, and 76 regions of zones, for a total of 111 geographic divisions. However, six of these zones (see Table B.1 in the Appendix B) are each formed by a single region, resulting in a total of 105 unique nodes in the hierarchy. The purpose of travel is divided into four categories: holiday (Hol), visiting friends and relatives (Vis), business (Bus), and other (Oth). According to Di Fonzo & Girolimetto (2022*b*), 24 nodes should not be considered, resulting in an unbalanced hierarchy of 525 unique nodes, referred to as VN525, instead of the theoretical 555. The dataset includes the 304 most disaggregated series, which can be aggregated into 221 upper time series for a total of 525 variables. Table 9 omits duplicated entries and updates

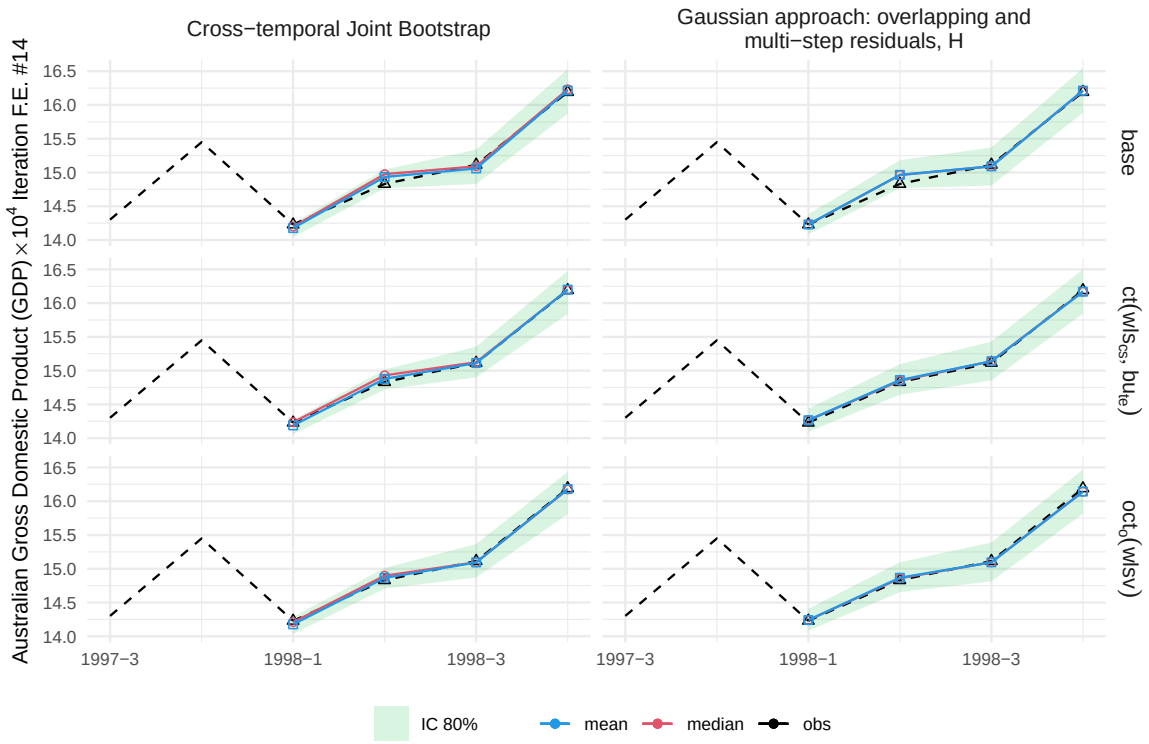


Figure 12: *Base (first row) and reconciled (second and third rows) forecasts for Australian GDP at the second iteration of our forecasting experiment for the bootstrap (first column) and Gaussian (second column) approaches. The green represents the 80% confidence interval. The black line and triangle indicate the observed true values, the red line represent the median value, and the blue line represent the mean value.*

the information in Table 7 from Wickramasuriya et al. (2019). This data can be temporally aggregated into 2, 3, 4, 6, or 12 months ($\mathcal{K} = \{1, 2, 3, 4, 6, 12\}$).

The forecasting experiment involves using a recursive training sample with an expanding window length. The process begins by using the first 10 years of data, from January 1998 to December 2008, to make forecasts for the entire following year (2009). Then, the training set is increased by one month and used to forecast the next 12 months, repeating this process until the last training set is used (January 1998 to December 2015). This results in a total of 85 different forecast origins, with each one being replicated in the forecasting experiment. Additionally, forecasts are computed up to 6 steps ahead for time series aggregated over 2 months, up to 4 steps ahead for those aggregated over 3 months, up to 3 steps ahead for those aggregated over 4 months, up to 2 steps ahead for those aggregated over 6 months, and one step ahead for those aggregated over 12 months. ETS models (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).

8.1 Base and reconciled samples

To calculate the base forecasts sample, we use the cross-temporal joint bootstrap method described in Section 3.1 for the nonparametric framework. In the Gaussian scenario, we

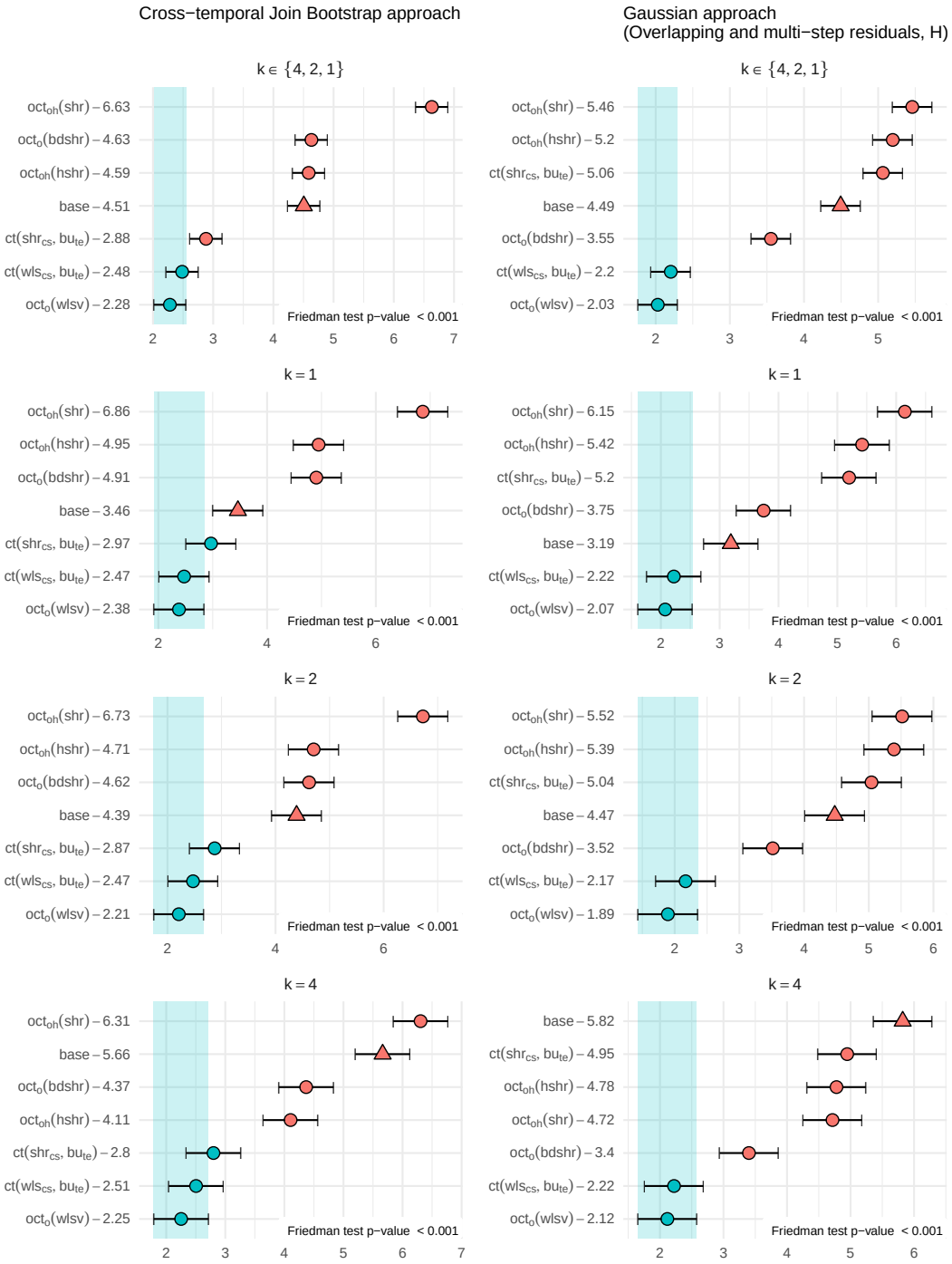


Figure 13: *"Multiple Comparison with the Best" (MCB) Nemenyi test at different temporal aggregation for the non-parametric bootstrap approach and the Gaussian approach using overlapping and multi-step residuals, with the covariance matrix calculated from the high frequency time series (H). In each panel, the Friedman test 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 were indicated if the intervals of two forecast reconciliation procedures did not overlap. Thus, approaches that did not overlap with the blue interval were considered significantly worse than the best, and vice versa.*

	Number of series		
	G.D.	**P.O.T**	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 9: *Grouped time series for Australian tourism flows.*

utilize all four covariance matrices outlined in Section 4 with $\lambda = 0$ (sample covariance matrix): the *Global* (G), *High frequency* (H), *Bottom time series* (B) and *High frequency Bottom time series* (HB) sample covariance matrices. To compute these matrices, we use multi-step residuals to accurately assess the temporal and cross-sectional relationships, but do not include overlapping, as we are unable to correctly determine the residuals for the overlapping series using ETS models (see Section 5.2).

To reconciled the 1000 base forecast samples, we have implemented the following approaches:

- $\mathbf{ct}(bu)$ cross-temporal bottom up (see Section 2.2);
- $\mathbf{ct}(\cdot, bu_{te})$ partly bottom-up (Section 2.2) starting from cross-sectionally reconciled forecasts using the *shr* approach (Table 1);
- $\mathbf{ct}(\cdot, bu_{cs})$ partly bottom-up (Section 2.2) starting from temporally reconciled forecasts using the *wlsv* approach (Table 1);
- $\mathbf{oct}(\cdot)$ optimal cross-temporal reconciliation for the *ols*, *struc*, *wlsv* and *bdshr* approach (see Section 2.1). When necessary (*wlsv* and *bdshr*), one-step residuals were used.;
- $\mathbf{oct}_h(\cdot)$ optimal cross-temporal reconciliation with multi-step residuals (see Section 5.1) for the shrinkage approaches presented in Section 4: *shr* stand for *Global shrinkage*, *hshr* for *High frequency shrinkage*, *bshr* for *bottom time series shrinkage*, *hbshr* for *High frequency bottom time series shrinkage*.

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 forecasts are naturally non negative (ETS model with the log-transformation), while this is not true for the Gaussian approach. In this case, to satisfy the constraint, we simply set negative values to zero. For the cross-temporal reconciliation, Di Fonzo & Girolimetto (2023*a*) offer two potential solutions: a state-of-the-art numerical optimization procedure like `osqp` (Stellato et al. 2022, 2020) and a simple heuristic strategy called set-negative-to-zero (sntz). In the sntz approach, possible negative values of the reconciled high frequency bottom time series forecasts without any non-negativity constraints are set to zero and finally a cross-temporal reconciliation bottom-up (see Section 2.2) is used to obtain the complete

cross-temporal forecasts. Di Fonzo & Girolimetto (2023*b*) found that both methods produced similar quality forecasts for the photovoltaic power generation dataset, but the optimization method required much more time and effort compared to the heuristic method. 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 normal.

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. These were introduced and discussed in Sections 6.3 and 7.1.

8.2 Results

The skill scores for CRPS and ES are shown in Tables 10 and 11, respectively. These tables are divided by different temporal levels and each column uses a different approach to calculate the base forecasts, referred to as "base". The different reconciliation approaches are indicated by the rows. The bootstrap method was used as a benchmark to calculate the skill scores. The bootstrap method indices were used as a benchmark.

Looking at the base rows of each table, base forecasts using a Gaussian approach are better in terms of both CRPS and ES compared to the bootstrap approach (our 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. This was confirmed also by the results in the GDP application. Bootstrap forecasts suffer from the limited number of residuals available and are unable to provide a detailed understanding of the distribution, leading 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 not big differences. For this reason, using only the high frequency bottom time series can be useful to estimate fewer parameters and reduce the initial dimensionality.

In the Gaussian case, bottom up $\mathrm{ct}(bu)$ and partly bottom-up techniques like $\mathrm{ct}(shr_{cs}, bu_{te})$ and $\mathrm{ct}(wlsv_{te}, bu_{cs})$ 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 ($k \in \{12, 6, 4, 3\}$). Overall, $\mathrm{oct}(bdshr)$ in terms of CRPS is almost always the best. The shrinkage approach $\mathrm{oct}_h(hshr)$ has good performance in the bootstrap case: it is competitive with $\mathrm{oct}(bdshr)$ in lower temporal level ($k \in \{1, 2\}$) and it is able to improve for $k \geq 3$. In terms of ES, $\mathrm{oct}(bdshr)$ still competitive, although it is not always the approach with the best improvement. In this case, approaches that attempt to establish a temporal and cross-sectional relationship, such as $\mathrm{oct}_h(hbshr)$ and $\mathrm{oct}_h(bshr)$, are better. It is also worth noting that techniques that don't make use of residuals like $\mathrm{oct}(ols)$ and $\mathrm{oct}(struc)$ prove to be competitive by consistently improving the base forecasts in terms of CRPS and ES.

Reconciliation approach	Base forecasts' sample approach									
	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
			$k=2$					$k=3$		
base	1.000	0.970	0.969	0.970	0.971	1.000	0.971	0.971	0.972	0.973
ct(bu)	1.189	0.999	0.999	0.999	0.999	1.273	1.010	1.010	1.010	1.010
ct(shr_{cs}, bu_{te})	1.015	0.972	0.970	0.972	0.970	1.041	0.977	0.974	0.977	0.974
ct($wlsv_{te}, bu_{cs}$)	1.016	0.971	0.971	0.970	0.970	1.046	0.976	0.976	0.974	0.974
oct(ols)	0.992	0.991	0.991	0.990	0.991	0.994	0.992	0.993	0.991	0.992
oct($struc$)	0.982	0.966	0.965	0.965	0.965	0.986	0.967	0.966	0.966	0.965
oct($wlsv$)	0.972	0.961	0.960	0.960	0.960	0.983	0.963	0.962	0.962	0.962
oct($bdshr$)	**0.964**	**0.958**	**0.957**	**0.956**	**0.956**	0.972	**0.960**	**0.958**	**0.957**	**0.957**
oct$_h$($hbshr$)	0.992	1.013	1.015	1.012	1.015	0.994	1.019	1.021	1.018	1.020
oct$_h$($bshr$)	0.997	1.015	1.018	1.013	1.017	0.999	1.021	1.022	1.018	1.022
oct$_h$($hshr$)	0.965	0.987	0.987	0.986	0.987	**0.971**	0.994	0.994	0.992	0.993
oct$_h$(shr)	1.005	0.986	0.978	0.976	0.975	1.009	0.986	0.978	0.976	0.976
			$k=4$					$k=6$		
base	1.000	0.973	0.973	0.974	0.975	1.000	0.976	0.976	0.978	0.978
ct(bu)	1.340	1.016	1.015	1.015	1.015	1.450	1.023	1.023	1.023	1.023
ct(shr_{cs}, bu_{te})	1.061	0.978	0.973	0.978	0.973	1.094	0.978	0.972	0.978	0.972
ct($wlsv_{te}, bu_{cs}$)	1.068	0.977	0.977	0.974	0.974	1.103	0.977	0.977	0.974	0.974
oct(ols)	0.993	0.991	0.992	0.990	0.990	0.989	0.989	0.989	0.987	0.986
oct($struc$)	0.986	0.965	0.964	0.964	0.963	0.986	0.961	0.960	0.959	0.957
oct($wlsv$)	0.990	0.962	0.961	0.961	0.960	1.001	0.960	0.959	0.958	0.957
oct($bdshr$)	0.977	**0.959**	**0.956**	**0.955**	**0.954**	0.985	**0.956**	**0.953**	**0.950**	**0.948**
oct$_h$($hbshr$)	0.993	1.021	1.023	1.019	1.021	0.989	1.024	1.026	1.022	1.022
oct$_h$($bshr$)	0.997	1.022	1.022	1.019	1.022	0.994	1.022	1.022	1.020	1.022
oct$_h$($hshr$)	**0.973**	0.996	0.997	0.994	0.995	**0.976**	1.000	1.001	0.996	0.997
oct$_h$(shr)	1.009	0.984	0.976	0.973	0.973	1.010	0.978	0.970	0.967	0.967
			$k=12$							
base	1.000	0.968	0.967	0.969	0.969					
ct(bu)	1.675	1.038	1.037	1.037	1.038					
ct(shr_{cs}, bu_{te})	1.163	0.977	0.965	0.977	0.965					
ct($wlsv_{te}, bu_{cs}$)	1.174	0.978	0.978	0.971	0.971					
oct(ols)	0.982	0.982	0.983	0.980	0.975					
oct($struc$)	0.982	0.951	0.949	0.947	0.943					
oct($wlsv$)	1.025	0.954	0.953	0.949	0.947					
oct($bdshr$)	1.002	**0.950**	**0.944**	**0.939**	**0.935**					
oct$_h$($hbshr$)	0.982	1.027	1.029	1.024	1.021					
oct$_h$($bshr$)	0.987	1.024	1.021	1.021	1.019					
oct$_h$($hshr$)	**0.978**	1.003	1.005	0.996	0.997					
oct$_h$(shr)	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 10: *CRPS skill score defined in equation* (17) *and* (18) *for VN525 dataset. The smaller this value, the more accurate the forecast. Approaches that performed worse than the benchmark model (Bootstrap base forecasts) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The notation used to refer to the reconciliation and base forecast samples is explained in Section 8.1.*

Reconciliation approach	Base forecasts' sample approach									
	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**
			$k=2$					$k=3$		
base	1.000	0.958	0.954	0.956	0.953	1.000	0.961	0.958	0.960	0.955
ct(bu)	2.176	1.001	1.001	1.001	1.001	2.428	0.998	0.997	0.997	0.997
ct(shr_{cs}, bu_{te})	1.192	0.927	0.921	0.927	0.921	1.245	0.911	0.904	0.911	0.904
ct($wlsv_{te}, bu_{cs}$)	1.400	0.992	0.992	0.988	0.988	1.500	0.991	0.991	0.986	0.987
oct(ols)	0.985	0.935	0.932	0.934	0.930	0.976	0.918	0.915	0.917	0.912
oct($struc$)	1.075	0.949	0.947	0.948	0.944	1.096	0.939	0.936	0.938	0.933
oct($wlsv$)	1.110	0.960	0.958	0.958	0.955	1.142	0.953	0.949	0.951	0.946
oct($bdshr$)	1.045	0.938	0.933	0.933	0.929	1.060	0.926	0.920	0.921	0.915
oct$_h$($hbshr$)	0.984	0.933	0.931	0.933	0.928	0.975	0.915	0.912	0.915	0.909
oct$_h$($bshr$)	**0.967**	**0.917**	**0.916**	0.913	0.908	**0.954**	**0.895**	**0.895**	**0.892**	**0.887**
oct$_h$($hshr$)	1.073	0.962	0.959	0.963	0.956	1.093	0.955	0.951	0.956	0.949
oct$_h$(shr)	1.064	0.933	0.916	**0.913**	**0.904**	1.082	0.923	0.903	0.900	0.890
			$k=4$					$k=6$		
base	1.000	0.960	0.960	0.962	0.956	1.000	0.961	0.959	0.964	0.956
ct(bu)	2.585	0.996	0.996	0.995	0.996	2.849	1.004	1.003	1.003	1.004
ct(shr_{cs}, bu_{te})	1.277	0.898	0.890	0.899	0.891	1.339	0.882	0.873	0.883	0.874
ct($wlsv_{te}, bu_{cs}$)	1.559	0.990	0.990	0.984	0.985	1.662	0.997	0.997	0.991	0.992
oct(ols)	0.966	0.905	0.902	0.904	0.899	0.962	0.889	0.887	0.890	0.885
oct($struc$)	1.106	0.930	0.927	0.928	0.924	1.132	0.923	0.919	0.922	0.916
oct($wlsv$)	1.157	0.947	0.943	0.945	0.939	1.192	0.942	0.937	0.941	0.934
oct($bdshr$)	1.065	0.917	0.909	0.910	0.903	1.084	0.907	0.897	0.898	0.890
oct$_h$($hbshr$)	0.967	0.901	0.898	0.900	0.895	0.964	0.882	0.880	0.883	0.877
oct$_h$($bshr$)	**0.943**	**0.879**	**0.878**	**0.876**	**0.871**	**0.932**	**0.856**	**0.855**	**0.851**	**0.848**
oct$_h$($hshr$)	1.101	0.949	0.944	0.949	0.941	1.126	0.945	0.939	0.945	0.936
oct$_h$(shr)	1.089	0.915	0.893	0.890	0.878	1.107	0.899	0.875	0.871	0.858
			$k=12$							
base	1.000	0.942	0.947	0.951	0.937					
ct(bu)	2.990	0.922	0.921	0.923	0.923					
ct(shr_{cs}, bu_{te})	1.326	0.779	0.767	0.777	0.766					
ct($wlsv_{te}, bu_{cs}$)	1.679	0.917	0.917	0.906	0.908					
oct(ols)	0.872	0.783	0.784	0.783	0.779					
oct($struc$)	1.077	0.826	0.822	0.823	0.818					
oct($wlsv$)	1.149	0.851	0.845	0.847	0.840					
oct($bdshr$)	1.021	0.808	0.796	0.796	0.787					
oct$_h$($hbshr$)	0.872	0.775	0.772	0.772	0.770					
oct$_h$($bshr$)	**0.833**	**0.741**	**0.741**	**0.737**	**0.735**					
oct$_h$($hshr$)	1.066	0.851	0.846	0.848	0.838					
oct$_h$(shr)	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 11: *ES skill score defined in equation* (17) *and* (19) *for VN525 dataset. The smaller this value, the more accurate the forecast. Approaches that performed worse than the benchmark model (Bootstrap base forecasts) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The notation used to refer to the reconciliation and base forecast samples is explained in Section 8.1.*

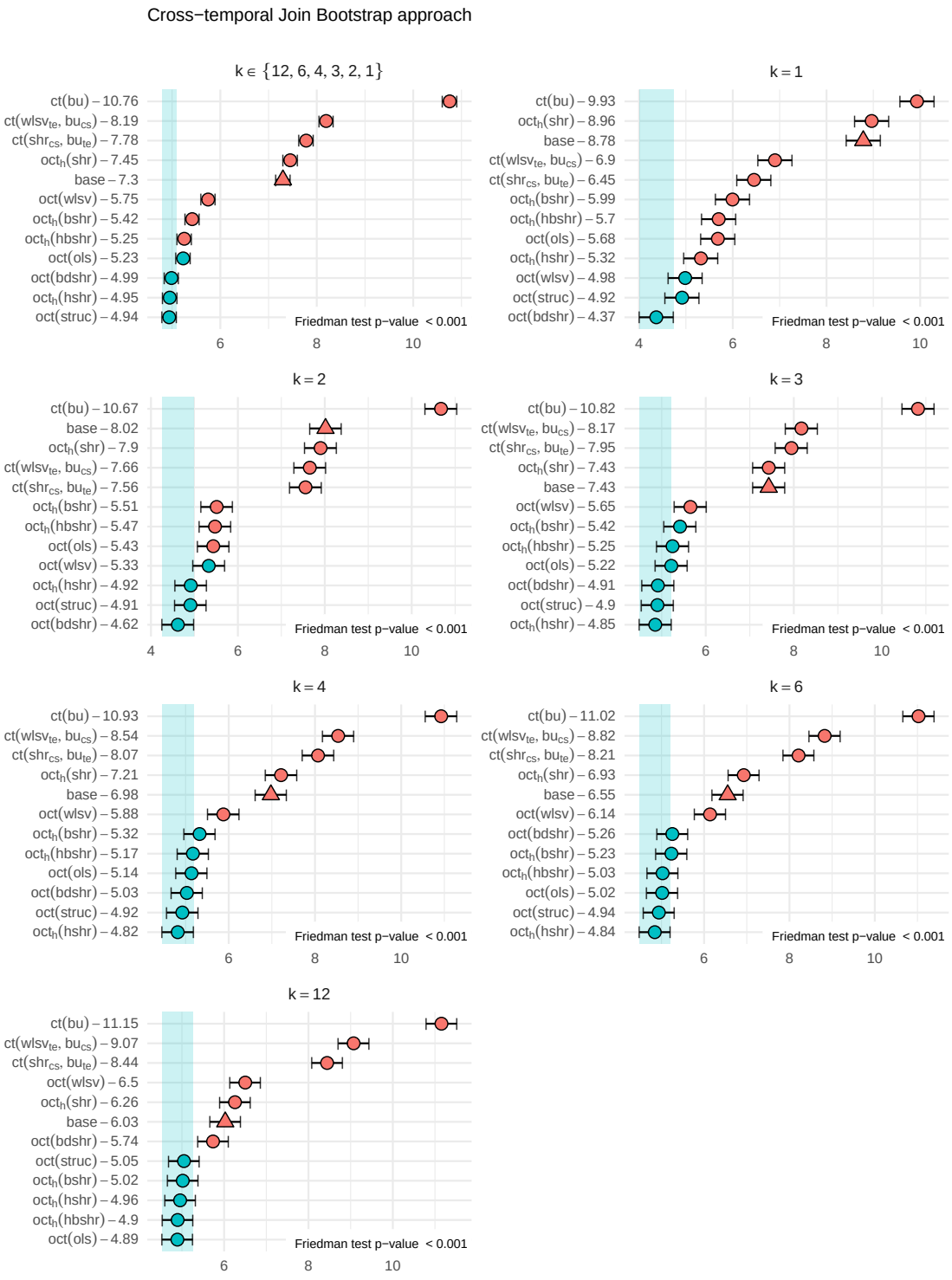


Figure 14: *"Multiple Comparison with the Best" (MCB) Nemenyi test at different temporal aggregation for the non-parametric bootstrap approach. In each panel, the Friedman test is reported in the lower right corner. The mean rank of each approach is shown to the right of its name. Statistical differences in performance were indicated if the intervals of two forecast reconciliation procedures did not overlap. Thus, approaches that did not overlap with the blue interval were considered significantly worse than the best, and vice versa.*

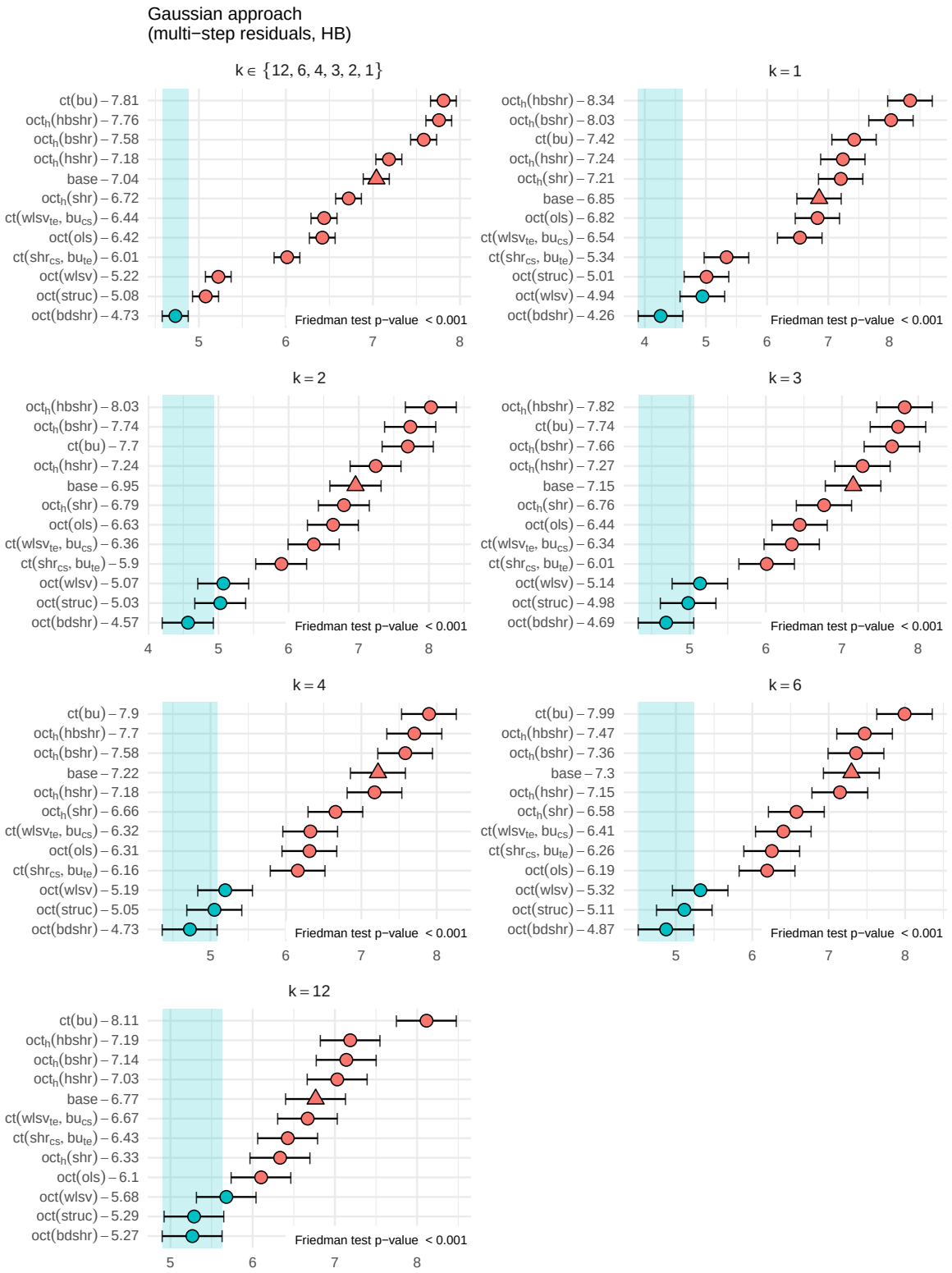


Figure 15: *"Multiple Comparison with the Best" (MCB) Nemenyi test at different temporal aggregation for the Gaussian approach using multi-step residuals, with the covariance matrix calculated from high frequency bottom time series (HB). In each panel, the Friedman test is reported in the lower right corner. The mean rank of each approach is shown to the right of its name. Statistical differences in performance were indicated if the intervals of two forecast reconciliation procedures did not overlap. Thus, approaches that did not overlap with the blue interval were considered significantly worse than the best, and vice versa.*

Figures 14 and 15 show the MCB using the CRPS for the non-parametric bootstrap approach and the Gaussian approach using multi-step residuals (HB). 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 procedure, oct$(bdshr)$, oct$(hshr)$ and oct$(struc)$ that are almost always in the group of the best approaches (underlying with 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.

9 Conclusion

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

Moreover, we analyzed the use of residuals, showing that one-step residuals fail to capture the temporal structure, and we proposed multi-step residuals that can fully capture the full cross-temporal relationships. When dealing with covariance matrices (due to the high-dimension), we proposed four alternative forms to reduce the number of parameters to be estimated. The overlapping residuals address the high-dimension problem by increasing the number of available residuals; this idea requires further investigation in future works. We presented a simple simulation to clarify what happens in cross-temporal cases using different types of residuals.

Finally, we performed two empirical analysis on two classical data sets commonly used in forecast reconciliation research: the Australian GDP, and the Australian Tourism Demand. We found that optimal cross-temporal reconciliation techniques are significantly better than base forecasts. We also compared them with partly bottom-up techniques that use uni-dimensional reconciliations (either cross-sectional or temporal) and found that they perform better, especially at higher levels of temporal aggregation. All the reconciliation methods proposed are available in R package `FoReco` (Girolimetto & Di Fonzo 2022). In conclusion, reconciliation approaches can play an important role to improve the accuracy of forecasts in a probabilistic framework.

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.

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. (2022), *FoReco: Point Forecast Reconciliation*. R package v0.2.5. **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.

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*

Wickramasuriya, S. L. (2021), 'Probabilistic forecast reconciliation under the Gaussian framework', *arXiv:2103.11128* .

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 The AR(2) processes' cross-temporal covariance matrix

Note that,

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

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_2\left(\mathbf{0}_{(2\times 1)}, \boldsymbol{\Omega}_{cs}\right)$. 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

$$\begin{array}{ll} Y_{i,T+1} &= \phi_{i,1} Y_{i,T} + \phi_{i,2} Y_{i,T-1} + \varepsilon_{i,T+1} \\ Y_{i,T+2} &= \phi_{i,1} Y_{i,T+1} + \phi_{i,2} Y_{i,T} + \varepsilon_{i,T+2} \end{array} \quad \text{and} \quad \begin{array}{ll} \widetilde{Y}_{i,T+1} &= \phi_{i,1} Y_{i,T} + \phi_{i,2} Y_{i,T-1} \\ \widetilde{Y}_{i,T+2} &= \phi_{i,1} \widetilde{Y}_{i,T+1} + \phi_{i,2} Y_{i,T} \end{array},$$

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,

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

and

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

B Geographic divisions of Australia

Table B.1: *Geographic divisions of Australia in States, Zones e Regions. Zones formed by a single region have been highlighted in italics.*

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

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