Cross-temporal Probabilistic Forecast Reconciliation

Daniele Girolimetto*
Department of Statistical Sciences, University of Padova
and
George Athanasopoulos
Department of Econometrics and Business Statistics, Monash University
and
Tommaso Di Fonzo
Department of Statistical Sciences, University of Padova
and
Rob J Hyndman
Department of Econometrics and Business Statistics, Monash University

April 21, 2023

Abstract

Forecast reconciliation is a post-forecasting process that involves transforming 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 current state-of-the-art cross-sectional probabilistic forecast reconciliation approach to encompass a cross-temporal framework, where temporal constraints are also applied. Our proposed methodology employs both parametric Gaussian and non-parametric bootstrap approaches to draw samples from an incoherent cross-temporal distribution. To improve the estimation of the forecast error covariance matrix, we propose using multi-step residuals, especially in the time dimension where the usual one-step residuals fail. To address high-dimensionality issues, we present four alternatives for the covariance matrix, where we exploit the two-fold nature (cross-sectional and temporal) of the cross-temporal structure, and introduce the idea of overlapping residuals. We assess the effectiveness of the proposed cross-temporal reconciliation approaches through a simulation study that investigates their theoretical and empirical properties and two forecasting experiments, using the Australian GDP and the Australian Tourism Demand datasets. For both applications, the optimal cross-temporal reconciliation approaches significantly outperform the incoherent base forecasts in terms of the Continuous Ranked Probability Score and the Energy Score. Overall, the results highlight the potential of the proposed methods to improve the accuracy of probabilistic forecasts and to address the challenge of integrating disparate scenarios while coherently taking into account short-term operational, medium-term tactical, and long-term strategic planning.

Keywords: Forecasting, Forecast reconciliation, Linearly constrained multiple time series, GDP, Tourism flows

*Corresponding author. E-mail: daniele.girolimetto@phd.unipd.it

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). There are many fields where forecast reconciliation is useful, such as when forecasting demand in supply chains with product categories (Punia et al., 2020; Kourentzes and Athanasopoulos, 2021), electricity demand and power generation (Spiliotis et al., 2020; Ben Taieb et al., 2021), GDP and its components (Athanasopoulos et al., 2020), tourist flows across geographic regions and travel purpose (Kourentzes and Athanasopoulos, 2019), and more. Moreover, effective decision-making depends on the support of accurate and coherent forecasts, making the use of forecast reconciliation methods increasingly popular in recent years. One of the main reason for this is the complexity of modern organizations, where different departments or groups generate forecasts at different levels of aggregation, using varying approaches and datasets. This can lead to significant differences in the operating decisions, as each division may have a different view of what is optimal for their area of responsibility. Forecast reconciliation enables the integration of these disparate forecasts by sharing information and breaking down the barriers between silos. By combining the forecasts from all levels, it can lead to greater accuracy and reduce errors that might have otherwise occurred due to different assumptions or biases. Another benefit of forecast reconciliation is that it enables organizations to make more informed decisions by providing a global and detailed view. Rather than relying on a single forecast generated by one division, reconciled forecasts can take into account a broader range of perspectives, thus resulting in more accurate and robust decisions. Additionally, reconciled forecasts can help organizations identify areas where different departments may need to collaborate more closely or share information more effectively, leading to greater efficiencies and better outcomes. Overall, the growing popularity of forecast reconciliation methods reflects the need for organizations to manage complexity, share information, and make better-informed decisions in an increasingly complex and dynamic environment.

Temporal reconciliation is another important aspect of forecast reconciliation that can help organizations to better align their forecasting efforts. This approach consists in reconciling forecasts that are generated at different time horizons, such monthly, quarterly or annual. This is particularly relevant for organizations that need to plan and make decisions over varying horizons that may have different levels of importance. For example, a retail company may need to reconcile monthly forecasts of sales with quarterly forecasts of revenue to ensure that they are aligned and consistent. This can help the company to make more accurate decisions about inventory, staffing, and pricing, as well as to identify trends and patterns over time (Kourentzes, 2022).

Classical reconciliation methods addressed the issue of incoherent forecasts in a cross-sectional hierarchy by forecasting only one level and using these to generate forecasts for the remaining series. The bottom-up approach (Dunn et al., 1976) starts by generating forecasts at the most disaggregate level and summing these to arrive at the desired forecasts for aggregate levels. On the other hand, the top-down approach (Gross and 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. All of these approaches ignore useful information available at other levels (Pennings and van Dalen, 2017).

Consequently, hierarchical forecasting has significantly evolved to include modern least squares-based reconciliation techniques in the cross-sectional framework (Hyndman et al., 2011; Wickramasuriya et al., 2019; Panagiotelis et al., 2021), later extended to temporal hierarchies (Athanasopoulos et al., 2017; Nystrup et al., 2020). Obtaining coherent forecasts across both the cross-sectional and temporal dimensions (known as *cross-temporal coherence*) has been limited to sequential approaches that address each dimension separately (Kourentzes and Athanasopoulos, 2019; Yagli et al., 2019; Punia et al., 2020; Spiliotis et al., 2020). Recently, Di Fonzo and Girolimetto (2023a) 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 and Katzfuss, 2014). In the cross-sectional and temporal frameworks, there have been some developments towards probabilistic forecasting including Ben Taieb et al. (2017), Panamtash and Zhou (2018), Jeon et al. (2019), Yang (2020), Yagli et al. (2020), Ben Taieb et al. (2021), Corani et al. (2021), Corani et al.

(2022), Zambon et al. (2022) and Wickramasuriya (2023). Panagiotelis et al. (2023) made a significant contribution by formalizing cross-sectional probabilistic reconciliation using the geometric framework for point forecast reconciliation of Panagiotelis et al. (2021). They show how a reconciled forecast can be constructed from an arbitrary base forecast when its density is available and when only a sample can be drawn. They also show that in the case of elliptical distributions, the correct predictive distribution can be recovered via linear reconciliation, regardless of the base forecast location and scale parameters, and derive conditions for this to hold in the special case of reconciliation via projection.

In this paper, we extend cross-sectional probabilistic reconciliation to the cross-temporal case, working on issues related to the two-fold nature of this framework. First, we revise and develop the notation proposed by Di Fonzo and Girolimetto (2023a) to generalize the work of Panagiotelis et al. (2023). This allows us to move from cross-temporal point reconciliation to a probabilistic setting through the generalization of definitions and theorems well-established in the cross-sectional framework. Second, we propose effective and practical solutions to draw a sample from the base forecast distribution according to either a parametric approach that assumes Gaussianity or a non-parametric approach that bootstraps the base model residuals. Third, we propose some solutions to specific problems that arise when combining the cross-sectional and temporal dimensions. We propose using multi-step residuals to estimate the relationships between different forecast horizons when we deal with temporal levels, since one-step residuals are not suitable for this purpose. To solve high-dimensionality issues we introduce the idea of overlapping residuals and consider alternative forms for constructing the covariance matrix. Fourth, we propose new shrinkage procedures for reconciliation that aim to identify a feasible cross-temporal structure. The methodological contributions described in this paper are implemented in the `FoReco` package (Girolimetto and Di Fonzo, 2023) for R (R Core Team, 2022). Furthermore, the Appendix contains complementary materials on methodological and practical issues, and supplementary tables and graphs related to the empirical applications.

The remainder of the paper is structured as follows. In Section 2, we provide a unified notation for the cross-sectional, temporal and cross-temporal point reconciliation. We generalize the cross-sectional definitions and theorems developed by Panagiotelis et al. (2023) in Section 3, and propose both a parametric Gaussian and a non-parametric bootstrap

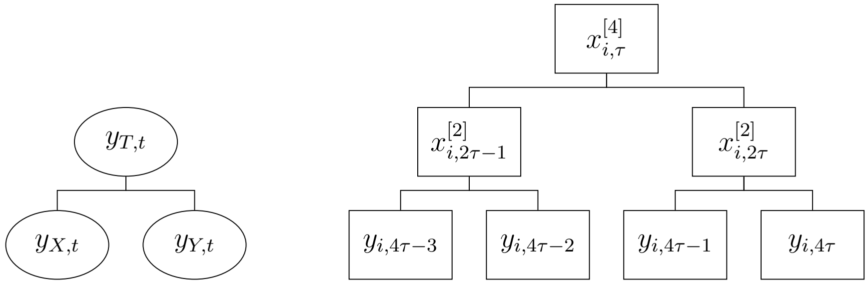

Figure 1: ***(a)*** *A simple two-level cross-sectional hierarchy for 3 time series with $n_a = 1$ and $n_b = 2$.* ***(b)*** *A temporal hierarchy for a quarterly series ($m = 4$ and $\mathcal{K} = \{4, 2, 1\}$).*

approach to draw a sample from the base forecast distribution. In Section 4, we analyze the structure of the cross-temporal covariance matrix, proposing four alternative forms, and propose shrinkage approaches for reconciliation. In addition, we explore cross-temporal residuals (overlapping and multi-step) looking at their advantages and limitations. Two empirical applications using the Australian GDP and the Australian Tourism Demand datasets are considered in Sections 5 and 6, respectively[1]. Finally, Section 7 presents conclusions and a future research agenda on this and other related topics.

2 Notation and definitions

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

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

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. Figure 1(a) shows the two-level hierarchical structure for three linearly constrained time series such that $y_{T,t} = y_{X,t} + y_{Y,t}$, $\forall t = 1, ..., T$. Now let $\boldsymbol{y}_t = [\boldsymbol{u}_t' \quad \boldsymbol{b}_t']'$, where $\boldsymbol{u}_t = [y_{1,t}, \dots, y_{n_a,t}]'$ is the n_a-

[1] A complete set of results is available at the GitHub repository `https://github.com/danigiro/ctprob`.

vector of upper levels time series and $\boldsymbol{b}_t = \begin{bmatrix} y_{(n_a+1),t} & \dots & y_{n,t} \end{bmatrix}'$ 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 and Girolimetto (2022c), we can always construct a zero-constraints cross-sectional matrix from the aggregation matrix, $\boldsymbol{C}_{cs} = \begin{bmatrix} \boldsymbol{I}_{n_a} & -\boldsymbol{A}_{cs} \end{bmatrix}$. Finally, the cross-sectional structural matrix is given by $\boldsymbol{S}_{cs} = \begin{bmatrix} \boldsymbol{A}_{cs} \\ \boldsymbol{I}_{n_b} \end{bmatrix}$, providing the structural representation (Hyndman et al., 2011) $\boldsymbol{y}_t = \boldsymbol{S}_{cs}\boldsymbol{b}_t$. Considering the hierarchical example in Figure 1(a), we have $\boldsymbol{A}_{cs} = [1 \quad 1]$, $\boldsymbol{C}_{cs} = [1 \quad -1 \quad -1]$. In general there is no reason for $\boldsymbol{u}_t$ to be restricted to simple sums of $\boldsymbol{b}_t$; therefore $\boldsymbol{A}_{cs} \in \mathbb{R}^{n_a \times n_b}$ may contain any real values, and not only 0s and 1s.

Considering now the temporal framework, we denote as $\mathcal{K} = \{k_p, k_{p-1}, \dots, k_2, k_1\}$ the set of p factors of m, in descending order, where $k_1 = 1$ and $k_p = m$ (Athanasopoulos et al., 2017). Given a factor k of m, and assuming that $T = Nm$ (where N is the length of the most temporally aggregated version of the series), we can construct a temporally aggregated version of the time series of a single variable $\{\boldsymbol{y}_{i,t}\}_{t=1,\dots,T}$, through the non-overlapping sums of its k successive values, which has a seasonal period equal to $M_k = \dfrac{m}{k}$: $x_{i,j}^{[k]} = \displaystyle\sum_{t=(j-1)k+1}^{jk} y_{i,t}$, where $j = 1, \dots, N_k$, $i = 1, \dots, n$, $N_k = \dfrac{T}{k}$ and $x_{i,j}^{[1]} = y_{i,t}$. Define τ as the observation index of the most aggregate level k_p. For a fixed temporal aggregation order $k \in \mathcal{K}$, we stack the observations in the column vector $\boldsymbol{x}_{i,\tau}^{[k]} = \begin{bmatrix} x_{i,M_k(\tau-1)+1}^{[k]} & x_{i,M_k(\tau-1)+2}^{[k]} & \dots & x_{i,M_k\tau}^{[k]} \end{bmatrix}'$, and obtain the vector for all the temporal aggregation orders $\boldsymbol{x}_{i,\tau} = \begin{bmatrix} x_{i,\tau}^{[k_p]} & \boldsymbol{x}_{i,\tau}^{[k_{p-1}]\prime} & \dots & \boldsymbol{x}_{i,\tau}^{[1]\prime} \end{bmatrix}'$, $\tau = 1, \dots, N$. The structural representation of the temporal hierarchy (Athanasopoulos et al., 2017) is then $\boldsymbol{x}_{i,\tau} = \boldsymbol{S}_{te}\boldsymbol{x}_{i,\tau}^{[1]}$, where $\boldsymbol{S}_{te} = \begin{bmatrix} \boldsymbol{A}_{te} \\ \boldsymbol{I}_m \end{bmatrix}$ is the $[(m + k^*) \times m]$ temporal structural matrix, $\boldsymbol{A}_{te} = \begin{bmatrix} \boldsymbol{1}_{k_p} & \boldsymbol{I}_{\frac{m}{k_{p-1}}} \otimes \boldsymbol{1}_{k_{p-1}} & \dots & \boldsymbol{I}_{\frac{m}{k_2}} \otimes \boldsymbol{1}_{k_2} \end{bmatrix}'$ is the $(k^* \times m)$ temporal aggregation matrix with $k^* = \displaystyle\sum_{k \in \mathcal{K} \setminus \{k_1\}} M_k$, and $\otimes$ is the Kronecker product. For each series $x_{i,\tau}$, $i = 1, \dots, n$, we have also the zero-constrained representation

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

where $\boldsymbol{C}_{te} = [\boldsymbol{I}_{k^*} \quad -\boldsymbol{A}_{te}]$ is the $[k^* \times (m + k^*)]$ zero constraints temporal matrix. Figure 1(b) shows the hierarchical representation of a quarterly time series, for which $m = 4$, $\mathcal{K} = \{4, 2, 1\}$.When we temporally aggregate each series, the cross-sectional constraints for the most temporally disaggregated series (1) hold for all the temporal aggregation orders such that $\boldsymbol{C}_{cs}\boldsymbol{x}_j^{[k]} = \boldsymbol{0}_{(n_a\times 1)}$, for $k \in \mathcal{K}$ and $j = 1, \dots, N_k$, where $\boldsymbol{x}_j^{[k]} = \begin{bmatrix}\boldsymbol{u}_j^{[k]\prime} & \boldsymbol{b}_j^{[k]\prime}\end{bmatrix}^\prime$ with $\boldsymbol{u}_j^{[k]} = \begin{bmatrix}x_{1,\,j}^{[k]} & \dots & x_{n_a,\,j}^{[k]}\end{bmatrix}^\prime$ is the n_a-vector of upper time series and $\boldsymbol{b}_j^{[k]} = \begin{bmatrix}x_{(n_a+1),\,j}^{[k]} & \dots & x_{n,\,j}^{[k]}\end{bmatrix}^\prime$ 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: $\boldsymbol{X}_\tau = \begin{bmatrix}\boldsymbol{x}_{1,\tau}^\prime & \dots & \boldsymbol{x}_{n,\tau}\end{bmatrix}^\prime = \begin{bmatrix}\boldsymbol{X}_\tau^{[k_p]} & \dots & \boldsymbol{X}_\tau^{[k_1]}\end{bmatrix}$ with $\boldsymbol{X}_\tau^{[k_p]} = \begin{bmatrix}\boldsymbol{U}_\tau^{[k]\prime} & \boldsymbol{B}_\tau^{[k]\prime}\end{bmatrix}^\prime$, where for any fixed k, $\boldsymbol{U}_\tau^{[k]}$ is the $(n_a \times N_k)$ matrix grouping the upper time series, $\boldsymbol{B}_\tau^{[k]}$ is the $(n_b \times N_k)$ matrix grouping the bottom time series. Further, $\boldsymbol{C}_{cs}\boldsymbol{X}_\tau = \boldsymbol{0}_{[n_a\times(m+k^*)]}$ and $\boldsymbol{C}_{te}\boldsymbol{X}_\tau^\prime = \boldsymbol{0}_{(k^*\times n)}$. We can consider the cross-temporal framework as a generalization of the cross-sectional and temporal frameworks, that simultaneously takes into account both types of constraints. The cross-sectional reconciliation approach proposed by Hyndman et al. (2011) can be obtained by assuming $m = 1$, while the temporal one (Athanasopoulos et al., 2017) is obtained when $n = 1$ (with $n_a = 0$ and $n_b = 1$).

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

$$\boldsymbol{C}_{ct} = \begin{bmatrix}\boldsymbol{C}_* \\ \boldsymbol{I}_n \otimes \boldsymbol{C}_{te}\end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{C}_{ct}\boldsymbol{x}_\tau = \boldsymbol{0}_{[(n_a m+nk^*)\times 1]} \quad \text{for} \quad \tau = 1, \dots, N, \tag{3}$$

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

$$\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te} \tag{4}$$

is the $[n(k^* + m) \times n_b m]$ cross-temporal summation matrix, $s : \mathbb{R}^{n_b m} \to \mathbb{R}^{n(m+k^*)}$ is the operator describing the pre-multiplication by $\boldsymbol{S}_{ct}$, and $\boldsymbol{b}_\tau^{[1]} = \mathrm{vec}(\boldsymbol{B}_\tau^{[1]\prime})$.

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 $\boldsymbol{S}_{ct}$.

2.1 Optimal point forecast reconciliation

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

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

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 $\widehat{\boldsymbol{X}}_h$, contains incoherent forecasts, such as $\boldsymbol{C}_{ct}\widehat{\boldsymbol{x}}_h \neq \boldsymbol{0}_{[(n_a m + nk^*)\times 1]}$ with $h = 1, \ldots, H$ and $\widehat{\boldsymbol{x}}_h = \text{vec}(\widehat{\boldsymbol{X}_h'})$. 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 $h = 1, \ldots, H$, the mapping ψ may be defined as a projection onto $\mathfrak{s}$ given by (Panagiotelis et al., 2021; Di Fonzo and Girolimetto, 2023a)

$$\widetilde{\boldsymbol{x}}_h = \psi(\widehat{\boldsymbol{x}}_h) = \boldsymbol{M}\widehat{\boldsymbol{x}}_h, \tag{5}$$

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 $\widetilde{\boldsymbol{x}}_h = \text{vec}(\widetilde{\boldsymbol{X}_h'})$. Wickramasuriya et al. (2019) showed that the minimum variance linear unbiased reconciled forecasts, satisfying the unbiased condition $\text{E}(\widetilde{\boldsymbol{x}}_h - \boldsymbol{x}_h) = 0$, has solution (5) when $\boldsymbol{\Omega}_{ct} = \text{Var}(\widetilde{\boldsymbol{x}}_h - \boldsymbol{x}_h)$.

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

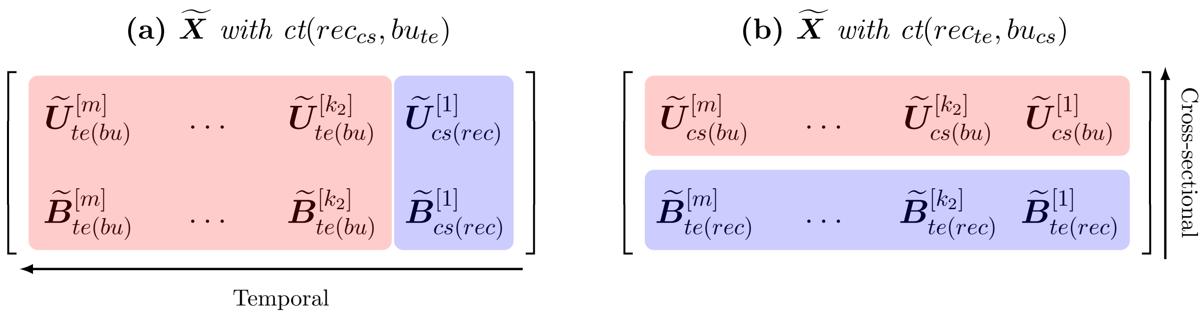

Figure 2: *A visual representation of partly bottom up starting from* (2a) *cross-sectionally reconciled forecasts for the temporal order 1 ($\widetilde{\boldsymbol{U}}^{[1]}$ and $\widetilde{\boldsymbol{B}}^{[1]}$) followed by temporal bottom-up, and* (2b) *temporally reconciled forecasts of the cross-sectional bottom time series ($\widetilde{\boldsymbol{B}}^{[k]}$, $k \in \mathcal{K}$) followed by cross-sectional bottom-up.*

leads to a solution equivalent to the cross-temporally reconciled forecasts in (5), given by

$$\widetilde{\boldsymbol{x}}_h = \psi\left(\widehat{\boldsymbol{x}}_h\right) = (s \circ g)\left(\widehat{\boldsymbol{x}}_h\right) = \boldsymbol{S}_{ct}\boldsymbol{G}\widehat{\boldsymbol{x}}_h, \tag{6}$$

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 $g : \mathbb{R}^{n(m+k^*)} \rightarrow \mathbb{R}^{n_b m}$ is a continuous function. In Appendix A we report some cross-sectional, temporal and cross-temporal approximations for the covariance matrix to be used in (5) and (6).

2.2 Cross-temporal bottom-up forecast reconciliation

The classic bottom-up approach (Dunn et al., 1976; Dangerfield and 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 and Girolimetto, 2022a, 2023b; Sanguri et al., 2022).

Figure 2 provides a visual representation of partly bottom-up in a two-step cross-temporal reconciliation approach. On the left (Figure 2a), we first compute the cross-sectionally reconciled forecasts at the highest frequency ($k = 1$), and then apply temporal bottom-up to obtain coherent cross-temporal forecasts. On the right (Figure 2b), we first compute temporally reconciled forecasts for the most disaggregated cross-sectional level, and then apply the cross-sectional bottom-up. We denote these two-step reconciliation approaches, respectively, as $\text{ct}(rec_{te}, bu_{cs})$, and $\text{ct}(rec_{cs}, bu_{te})$, where '$rec_{te}$' and '$rec_{cs}$' denote a forecast reconciliation approach in the temporal and cross-sectional dimensions and, 'bu_{cs}' and 'bu_{te}' denote using bottom-up in the cross-sectional and temporal dimensions, respectively. It is worth noting that the simple cross-temporal bottom-up approach corresponds to $\text{ct}(bu_{cs}, bu_{te}) = \text{ct}(bu_{te}, bu_{cs}) = \text{ct}(bu)$.

3 Probabilistic forecast reconciliation

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

Our aim is to extend these definitions to *cross-temporal coherent probabilistic forecasts* and *cross-temporal probabilistic forecast reconciliation*. Let $(\mathbb{R}^{n_b m}, \mathcal{F}_{\mathbb{R}^{n_b m}}, \nu)$ be a probability space for the bottom time series $\boldsymbol{b}_\tau^{[1]}$, where $\mathcal{F}_{\mathbb{R}^{n_b m}}$ is the Borel σ-algebra on $\mathbb{R}^{n_b m}$. Then a σ-algebra $\mathcal{F}_{\mathfrak{s}}$ can be constructed from the collection of sets $s(\mathcal{B})$ for all $\mathcal{B} \in \mathcal{F}_{\mathbb{R}^{n_b m}}$.

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

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

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

$$\tilde{\nu}(\mathcal{A}) = \hat{\nu}(\psi^{-1}(\mathcal{A})), \quad \forall \mathcal{A} \in \mathcal{F}_{\mathfrak{s}}, \tag{7}$$

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

The map ψ may be obtained as the composition $s \circ g$, as for the cross-temporal point reconciliation (6).

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 $\widetilde{\nu}$ defined in (7).

Proof. For any $\mathcal{A} \in \mathcal{F}_{\mathfrak{s}}$

$$\begin{aligned}
\Pr\left(\widehat{\boldsymbol{x}} \in \psi^{-1}(\mathcal{A})\right) &= \lim_{L\to\infty} \sum_{\ell=1}^{L} \mathbb{1}\left\{\widehat{\boldsymbol{x}}_\ell \in \psi^{-1}(\mathcal{A})\right\} \\
&= \lim_{L\to\infty} \sum_{\ell=1}^{L} \mathbb{1}\left\{\psi\left(\widehat{\boldsymbol{x}}_\ell\right) \in \mathcal{A}\right\} = \Pr\left(\widetilde{\boldsymbol{x}} \in \mathcal{A}\right)
\end{aligned}$$

$\square$

Theorem 3.1 is the cross-temporal extension of Theorem 4.5 in Panagiotelis et al. (2023), valid only for the cross-sectional case. It means that a sample from the reconciled distribution can be obtained by reconciling each member of a sample from the incoherent distribution. With this result, we can separate the mechanism used to generate the base forecasts samples from the reconciliation phase.

3.1 Parametric framework: Gaussian reconciliation

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

$$\widetilde{\boldsymbol{x}} = \boldsymbol{M}\widehat{\boldsymbol{x}} \quad \text{and} \quad \widetilde{\boldsymbol{\Omega}} = \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}', \tag{8}$$

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

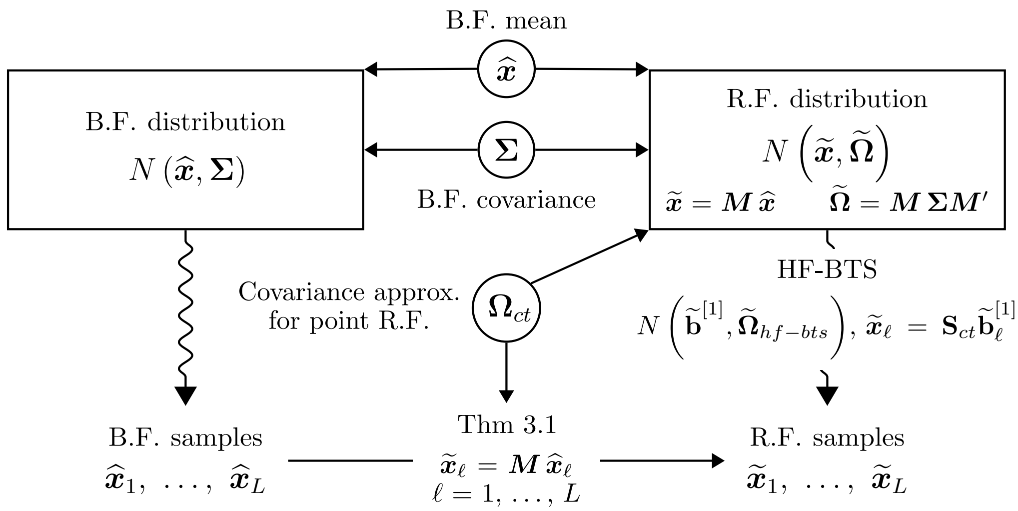

Figure 3: *Visual description of cross-temporal forecast reconciliation in the Gaussian framework, as described in Section 3.1. The acronyms R.F and B.F. stand for Reconciled and Base Forecasts, respectively. HF-BTS stands for High Frequency Bottom Time Series.*

where $\boldsymbol{M}$ is the projection matrix defined in (5). Note that if we assume that $\boldsymbol{\Sigma} = \boldsymbol{\Omega}_{ct}$, then the covariance matrix in (8) simplifies to $\widetilde{\boldsymbol{\Omega}} = \boldsymbol{M}\boldsymbol{\Omega}_{ct}$. In the cross-temporal case, sensibly estimating the covariance matrix $\boldsymbol{\Sigma}$ can be difficult because we need to simultaneously consider both the temporal and cross-sectional structures. This requires many parameters to be estimated, which can be challenging in practice. Additionally, naively using one-step residuals to estimate the cross-temporal correlation structure can lead to an inappropriate estimate of the covariance matrix[3]. These challenges will be explored in more depth in the following sections.

Focusing on the computational aspect[4], we can take several steps to reduce the time required to obtain simulations from the reconciled forecast distribution. For example when dealing with a genuine hierarchical structure, it is not necessary to simulate from a normal distribution with a defined covariance matrix for the entire structure. Instead, we can utilize the properties of elliptical distributions to simulate from the high frequency bottom time series and then obtain the complete simulation through the $\boldsymbol{S}_{ct}$ matrix. Furthermore, we do not need to calculate the reconciled mean and variance and generate a new sample if we already have a sample from the normal distribution of the base forecasts; we can simply apply the point forecast reconciliation (5) as outlined in Theorem 3.1. The relationships

[3]In particular, some temporal covariances are fixed to zero (see Appendix C for more details).

[4]We use two R packages to sample from a the base forecast gaussian distribution: `MASS` (Venables and Ripley, 2002) and `Rfast` (Papadakis et al., 2022) in Sections 5 and 6, respectively.

between base and reconciled forecast distributions and their respective simulations through Theorem 3.1 are depicted in Figure 3.

3.2 Non-parametric framework: bootstrap reconciliation

Analytical expressions for the base and reconciled forecast distributions are sometimes challenging to obtain. Furthermore parametric assumptions can be restrictive and unrealistic. We propose a procedure called *cross-temporal joint (block) bootstrap* (**ctjb**) to generate samples from the base forecast distributions that preserve cross-temporal relationships. This approach involves drawing samples of all series simultaneously 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 $(n \times N_k)$ matrix of the residuals for $k \in \mathcal{K}$. Figure 4 (on the left) provides a visualization of these matrices and how they are related to each other for the example in Figure 1. It is assumed that the residuals cover four years ($N = 4$): the green color corresponds to the first year, the blue to the second year, and so on. Further, let $\mathcal{M}_i$ be the model used to calculate the base forecasts and residuals for the i^{th} series. Assuming $H = 1$, τ is a random draw with replacement from $1, \dots, N$ and the ℓ^{th} bootstrap incoherent sample is $\widehat{\boldsymbol{x}}_{i,\ell}^{[k]} = f_i(\mathcal{M}_i, \widehat{\boldsymbol{e}}_i^{[k]})$, where $f_i(\cdot)$ depends on the fitted model $\mathcal{M}_i$. That is, $\widehat{\boldsymbol{x}}_{i,l}^{[k]}$ is a sample path simulated for the i^{th} series with error approximated by the corresponding block bootstrapped sample residual $\widehat{\boldsymbol{e}}_i^{[k]}$, the i^{th} row of

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

Figure 4 (on the right) shows $\widehat{\boldsymbol{E}}_{\tau}^{[k]}$ 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 for many forecasting models, making it

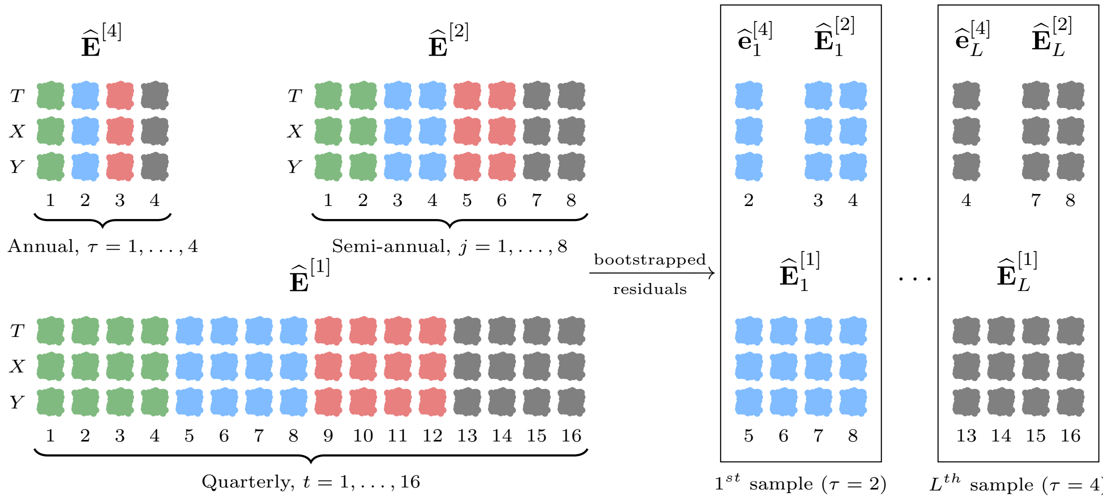

Figure 4: *Example of bootstrapped residuals for 3 linearly constrained quarterly time series (see Figure 1). On the left there are the residual matrices with 4 years of data ($N = 4$): the green, blue, red and black colors correspond, respectively, to years 1, 2, 3 and 4. On the right the bootstrapped residuals are represented.*

a practical and efficient tool. Furthermore, this approach is easily scalable in order to utilize multiple computing power simultaneously for each individual series. This can be especially useful when dealing with large datasets or when trying to speed up the analysis process.

4 Cross-temporal covariance matrix estimation

As the covariance matrix $\mathbf{\Omega}$ is unknown in practice, a natural estimate is the empirical sample covariance matrix of the base forecasts $\widehat{\mathbf{\Omega}}$. In this section, our focus will be exclusively on the cross-temporal framework., this means that we have to estimate $r = n(k^* + m)[n(k^* + m) - 1]/2$ different parameters. A possible solution to estimating many parameters when we have fewer observations than r, is to construct a shrinkage estimator (Efron, 1975; Efron and Morris, 1975, 1977), using a convex combination of $\widehat{\mathbf{\Omega}}$ and a diagonal target matrix $\widehat{\mathbf{\Omega}}_D = \widehat{\mathbf{\Omega}} \odot \boldsymbol{I}_{n(k^*+m)}$, such that $\widehat{\mathbf{\Omega}}_G = \lambda\widehat{\mathbf{\Omega}}_D + (1 - \lambda)\widehat{\mathbf{\Omega}}$, where $\lambda \in [0, 1]$ is the shrinkage intensity parameter that can be estimate using the unbiased estimator proposed by Ledoit and Wolf (2004) (see Schäfer and Strimmer, 2005). The linear combination involving these two matrices is referred to as *Global shrinkage* (G), where all off-diagonal elements are shrunk towards zero. $\widehat{\mathbf{\Omega}}_G$ corresponds to the matrix used by the reconciliation approach oct(shr) shown in Appendix A. However, shrinking all off-diagonal elements to zero, when we know that the covariance matrix has a cross-sectional and/or temporal structure, results in information loss. Therefore, we propose to estimate a smaller matrix, and to use the cross-sectional and/or temporal structure to obtain a better estimator for the covariance matrix of the entire system. Given that $\boldsymbol{S}_{ct} = \boldsymbol{S}_{cs} \otimes \boldsymbol{S}_{te}$, it is possible to express the actual covariance matrix in terms of three smaller matrices such that

$$\begin{aligned}\boldsymbol{\Omega} &= \boldsymbol{S}_{ct}\boldsymbol{\Omega}_{hf\text{-}bts}\boldsymbol{S}_{ct}' \\ &= (\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})\,\boldsymbol{\Omega}_{hf}\,(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})' \\ &= (\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})\,\boldsymbol{\Omega}_{bts}\,(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})',\end{aligned} \tag{9}$$

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

Therefore, we can apply the idea of "Stein-type shrinkage" (Efron and Morris, 1977) to $\boldsymbol{\Omega}_{hf\text{-}bts}$, $\boldsymbol{\Omega}_{hf}$ and $\boldsymbol{\Omega}_{bts}$ by using the corresponding empirical base forecasts residuals estimation. We obtain the following expressions (see Appendix B for details):

- *High frequency Bottom time series shrinkage matrix* (HB):
  $\widehat{\boldsymbol{\Omega}}_{HB} = \lambda \boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts,D}\boldsymbol{S}_{ct}' + (1-\lambda)\boldsymbol{S}_{ct}\widehat{\boldsymbol{\Omega}}_{hf\text{-}bts}\boldsymbol{S}_{ct}'$;
- *High frequency shrinkage matrix* (H):
  $\widehat{\boldsymbol{\Omega}}_{H} = \lambda(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})\widehat{\boldsymbol{\Omega}}_{hf,D}(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})' + (1-\lambda)(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})\widehat{\boldsymbol{\Omega}}_{hf}(\boldsymbol{I}_n \otimes \boldsymbol{S}_{te})'$;
- *Bottom time series shrinkage matrix* (B):
  $\widehat{\boldsymbol{\Omega}}_{B} = \lambda\,(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})\,\widehat{\boldsymbol{\Omega}}_{bts,D}\,(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})' + (1-\lambda)\,(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})\,\widehat{\boldsymbol{\Omega}}_{bts}\,(\boldsymbol{S}_{cs} \otimes \boldsymbol{I}_{m+k^*})'$,

where $\widehat{\boldsymbol{\Omega}}_{l,D} = \boldsymbol{I}_{n_b m} \odot \widehat{\boldsymbol{\Omega}}_j$, $l = \{hf\text{-}bts,\ hf,\ bts\}$, and λ is the shrinkage parameter.

Another important aspect is the number of parameters to be estimated through the residuals of the base forecasts. In Table 1 we report the number of different parameters for the two forecasting experiment: Australian GDP (see Section 5) and Australian Tourism Demand (see Section 6). In addition, we also calculate the percentage reductions in the number of parameters compared to the global approach. As we can see, G involves a considerably large number of parameters compared to other estimators. *HB* leads to the largest decrease of around 85%, whereas approaches H and B lie somewhere between G and *HB*. In general, as m and n increase (see Appendix B), using H requires the estimation of less parameters than B.

In the forecasting experiments that follow and in the simulation in Appendix C, we

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

Table 1: *Number of different parameters that need to be estimated for the Australian GDP (see Section 5) and the Australian Tourism Demand (see Section 6) forecasting experiments. The percentage reductions in the number of parameters compared to the global approach are reported in parentheses.*

closely analyze these different constructions with a dual purpose. In particular, we use the full covariance matrix ($\lambda = 0$) of the base forecasts to obtain base forecast samples of the linearly constrained time series under Gaussianity. We also use the shrinkage versions as approximations of the covariance matrix to be used for reconciliation. This will allow us to better understand the properties and abilities of each parameterization.

4.1 Multi-step residuals

Model residuals may be used to estimate the covariance matrix in cross-temporal forecast reconciliation. In time series analysis, it is common to use residuals corresponding to one-step ahead forecasts. However, due to the temporal dimension in our setting, residuals corresponding to different forecast horizons are required. Thus, we define *multi-step residuals* as $e^{[k]}_{i,h,j} = x^{[k]}_{i,j+h} - \widehat{x}^{[k]}_{i,j+h|j}$, where $i = 1, \dots, n$, $j = 1, \dots, N_k$ and $\widehat{x}^{[k]}_{i,j+h|t}$ is the h-step fitted value, calculated as the h-step-ahead forecast using data up to time j. In general, these residuals will be autocorrelated except when $h = 1$.

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

$\boldsymbol{e}_{i,h}^{[k]} = \begin{bmatrix} e_{i,h,1}^{[k]} & e_{i,h,2}^{[k]} & \dots & e_{i,h,N_k}^{[k]} \end{bmatrix}'$ with $h = 1, \dots, M_k$, can be organized in matrix form as

$$\boldsymbol{E}_i^{[k]} = \begin{bmatrix} e_{i,1,1}^{[k]} & e_{i,2,2}^{[k]} & \dots & e_{i,M_k,M_k}^{[k]} \\ \vdots & \vdots & & \vdots \\ e_{i,1,N_k-M_k+1}^{[k]} & e_{i,2,N_k-M_k+2}^{[k]} & \dots & e_{i,M_k,N_k}^{[k]} \end{bmatrix}.$$

Let $\boldsymbol{E}_i = \begin{bmatrix} \boldsymbol{E}_i^{[m]} & \boldsymbol{E}_i^{[k_p-1]} & \dots & \boldsymbol{E}_i^{[1]} \end{bmatrix}$. Then the $[N \times n(m + k^*)]$ cross-temporal residual matrix is given by $\boldsymbol{E} = \begin{bmatrix} \boldsymbol{E}_1 & \boldsymbol{E}_2 & \dots & \boldsymbol{E}_n \end{bmatrix}$.

To better understand the properties of the proposed alternatives, a simulation study was performed, and the results are shown in Appendix C. We have studied the effect of combining cross-sectional and temporal aggregations using a simple hierarchy, where the small size and nature of the data generating process make it possible to exactly calculate the true cross-temporal covariance structure, thus providing insights into the nature of the time series data involved in the forecast reconciliation process. We find that simulating base forecasts from multi-step residuals allows for a more accurate estimation of the covariance matrix and that reconciliation further improves the forecast accuracy.

4.2 Overlapping residuals

Another issue that arises in the case of cross-temporal reconciliation is the low number of available residuals, especially for the higher orders of temporal aggregation. A possible solution is to use residuals calculated using overlapping series by allowing the year to have a varying starting time. To better explain how to calculate overlapping residuals, assume we have a single series $\boldsymbol{y} = [y_1 \; y_2 \; y_3 \; \dots \; y_{T-1} \; y_T]'$. We can construct k non overlapping series such that $\boldsymbol{x}^{[k],s} = \left\{ x_j^{[k],s} \right\}_{j=1}^{N_k - s}$ where $x_j^{[k],s} = \sum\limits_{t=(j-1)k+s+1}^{jk-s} y_t$, with $s = 0, \dots, (k-1)$. For example, suppose we have a biannual series with $k = 2$ and $T = 6$, then we can construct two annual time series depending on which time is deemed the start of the year: $\boldsymbol{x}^{[2],0} = \begin{bmatrix} x_1^{[2],0} & x_2^{[2],0} & x_3^{[2],0} \end{bmatrix}' = \begin{bmatrix} y_1 + y_2 & y_3 + y_4 & y_5 + y_6 \end{bmatrix}'$ and $\boldsymbol{x}^{[2],1} = \begin{bmatrix} x_1^{[2],1} & x_2^{[2],1} \end{bmatrix}' = \begin{bmatrix} y_2 + y_3 & y_4 + y_5 \end{bmatrix}'$. To calculate overlapping residuals, we propose the following steps:

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

the parameters, and calculate the residuals.

The resulting residuals can be used to estimate the covariance matrix in cross-temporal forecast reconciliation. This increases the number of available residuals, particularly when working with higher frequency observations such as monthly or daily data. It is important to note that this approach assumes that the model used in step 1 is appropriate for all the different series $\boldsymbol{x}^{[k],s}$. Some seasonal models will not be appropriate as the seasonal pattern will be shifted for different values of s. However, this will not affect seasonal ARIMA models as the seasonality is defined in terms of lags which are unaffected by the value of s.

5 Forecasting Australian GDP

The Australian Quarterly National Accounts (QNA) dataset has been widely studied in the literature on forecast reconciliation (Athanasopoulos et al., 2020; Di Fonzo and Girolimetto, 2023a). Building on these results (see Appendix D.1), we now consider cross-temporally reconciled probabilistic forecasts.

We use univariate ARIMA models[5] to obtain quarterly base forecasts for the $n = 95$ QNA time series, spanning the period 1984:Q4 – 2018:Q1, defining GDP from both the Income and Expenditure sides. We perform a rolling forecast experiment with an expanding window: the first training sample spans the period 1984:Q4 to 1994:Q3, and the last ends in 2017:Q1, for a total of 91 forecast origins. For the temporal aggregation dimension we aggregate the quarterly data to both semi-annual and annual. We obtain 4-step, 2-step and 1-step ahead base forecasts respectively from the quarterly, semi-annual and annual frequencies, i.e., $\mathcal{K} = \{4, 2, 1\}$.

The base forecast samples in the Gaussian case are obtained using the sample covariance matrices with the *Global* (G) and *High frequency* (H) parameterization (Section 4), since it is not possible to identify a unique representation for the other cases[6]. We compare the results obtained using multi-step residuals with and without overlapping, in order to measure the benefit of obtaining overlapping residuals. In the non-parametric case, we

[5] We use the `auto.arima` function from the R package `forecast` (Hyndman et al., 2023).

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

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

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

use the cross-temporal joint bootstrap (ctjb) presented in Section 3.2. Finally, to reconcile the resulting (1000) base forecasts samples, we have applied the following techniques[7] (see Table 2): $\text{ct}(shr_{cs}, bu_{te})$, $\text{ct}(wls_{cs}, bu_{te})$, $\text{oct}_o(wlsv)$, $\text{oct}_o(bdshr)$, $\text{oct}_{oh}(shr)$ and $\text{oct}_{oh}(hshr)$.

The accuracy of the probabilistic forecasts is evaluated using the Continuous Ranked Probability Score (CRPS, Gneiting and Katzfuss, 2014), which 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 and Katzfuss, 2014), that is the CRPS extension to the multivariate case, to evaluate the forecasting accuracy for the whole system (Panagiotelis et al., 2023; Wickramasuriya, 2023). In particular, we consider the geometric mean of the

[7] In Appendix D.2, we show the results with shrunk covariance matrices. We also report the results obtained using one-step residuals in the reconciliation.

relative CRPS (Fleming and Wallace, 1986), and the relative ES:

$$\overline{\text{RelCRPS}}_{j,s}^{[k]} = \left(\prod_{i=1}^{n} \frac{CRPS_{i,j,s}^{[k]}}{CRPS_{i,0,0}^{[k]}}\right)^{\frac{1}{n}} \qquad \text{and} \qquad \text{RelES}_{j,s}^{[k]} = \frac{ES_{j,s}^{[k]}}{ES_{0,0}^{[k]}}, \tag{10}$$

where j denotes the reconciliation approach and s indicates the approach used to simulate the base forecasts. As a reference approach ($s = 0$ and $j = 0$), we consider the base forecasts produce by the Bootstrap approach. If we consider all the temporal aggregation orders (i.e. $\forall k \in \mathcal{K}$), the overall accuracy indices are given by, respectively,

$$\overline{\text{RelCRPS}}_{j,s} = \left(\prod_{\substack{i=1,\dots,n \\ k\in\mathcal{K}}} \frac{CRPS_{i,j,s}^{[k]}}{CRPS_{i,0,0}^{[k]}}\right)^{\frac{1}{n(k^*+m)}} \text{ and } \overline{\text{RelES}}_{j,s} = \left(\prod_{k\in\mathcal{K}} \frac{ES_{j,s}^{[k]}}{ES_{0,0}^{[k]}}\right)^{\frac{1}{(k^*+m)}}. \tag{11}$$

5.1 Results

Forecasting accuracy indices based on CRPS and ES are presented in Tables 3 and 4, respectively. As a benchmark approach, we use the base forecasts calculated using the bootstrap method. For base forecasts, we find that using a parametric approach with the normal distribution performs better than the non-parametric bootstrap approach. This is likely due to the limited number of residuals available for bootstrapping, which does not allow for sufficient exploration of the data. Directly specifying diagonal covariance matrices seems to be more effective than shrinking to a target covariance matrix. Among all the procedures, ct(wls_{cs}, bu_{te}) and oct$_o$($wlsv$) show the greatest relative gains. In contrast, oct$_{oh}$(shr) and oct_{oh}($hshr$) do not show much improvement. Furthermore, the greatest improvements are observed for higher temporal aggregation levels.

We utilize the non-parametric Friedman test and the post hoc "Multiple Comparison with the Best" (MCB) Nemenyi test (Koning et al., 2005; Kourentzes and Athanasopoulos, 2019; Makridakis et al., 2022) to determine if the forecasting performances of the different techniques are significantly different from one another. Figure 5 presents the MCB using the CRPS. The probabilistic forecasts from ct(wls_{cs}, bu_{te}) and oct$_o$($wlsv$) are significantly better than the base forecasts at any level of aggregation.

Overall, we find that using overlapping residuals almost always leads to a greater improvement in terms of both ES and CRPS. Forecasts at the most aggregated level (year)

Reconciliation approach	Generation of the base forecasts sample paths: ctjb	Gaussian approach* G_h	H_h	G_{oh}	H_{oh}	ctjb	Gaussian approach* 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 3: *$\overline{\text{RelCRPS}}$ defined in* (10) *and* (11) *for the Australian QNA dataset. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

Reconciliation approach	Generation of the base forecasts sample paths: ctjb	Gaussian approach* G_h	H_h	G_{oh}	H_{oh}	ctjb	Gaussian approach* 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 4: *ES ratio indices defined in* (10) *and* (11) *for the Australian QNA dataset. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

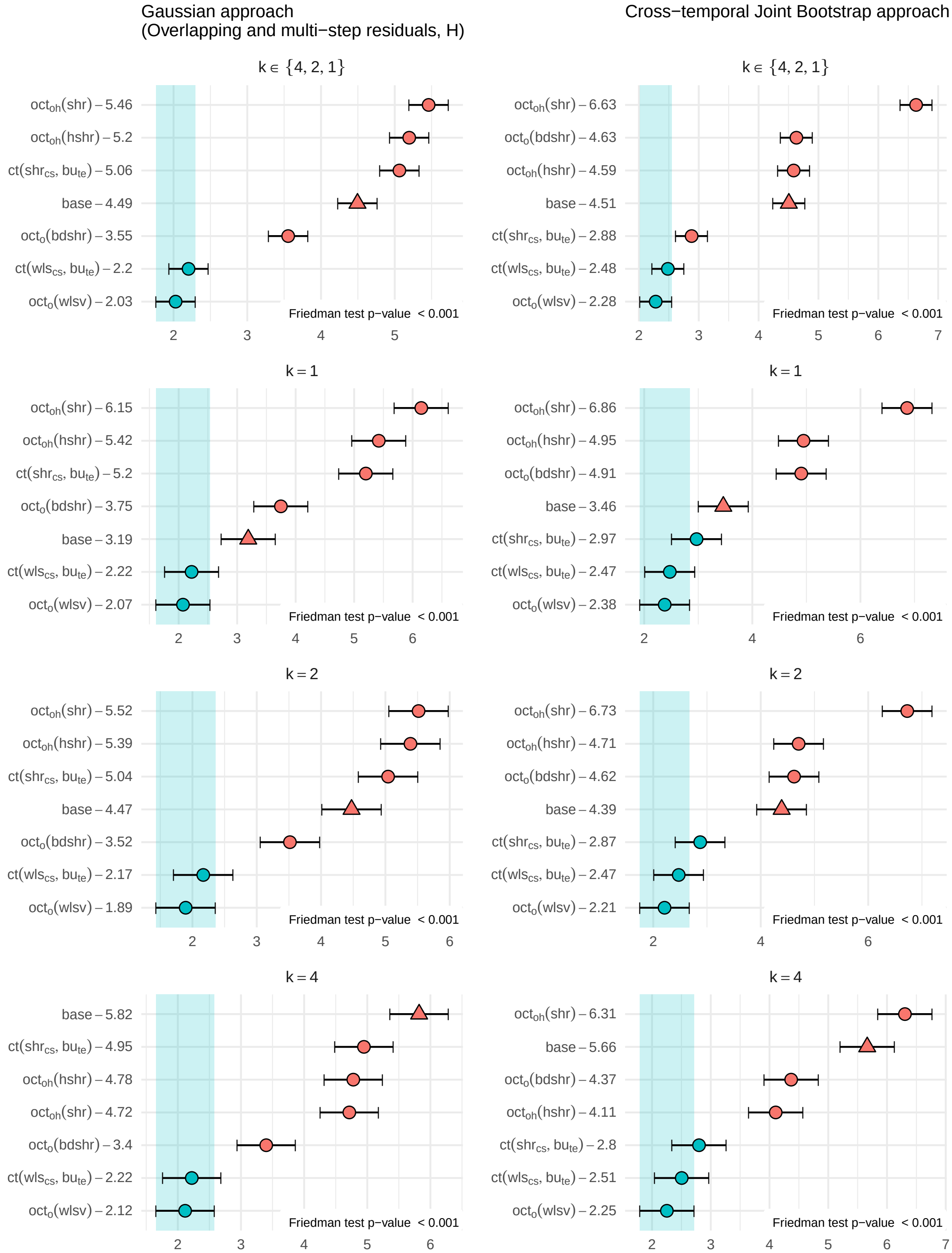

Figure 5: *MCB Nemenyi test for the Australian QNA dataset using CRPS at different temporal aggregation levels for the Gaussian (using overlapping and multi-step residuals, H) and the non-parametric bootstrap approaches. In each panel, the Friedman test p-value is reported in the lower right corner. The mean rank of each approach is shown to the right of its name. Statistical differences in performance are indicated if the intervals of two forecast reconciliation procedures do not overlap. Thus, approaches that do not overlap with the blue interval are considered significantly worse than the best, and vice-versa.*

seem to benefit the most from reconciliation, and using one-step overlapping residuals appears to be sufficient to improve forecasts if the generation of the base forecasts sample paths takes into account the multi-step structure.

6 Forecasting Australian Tourism Demand

The Australian Tourism Demand dataset (Wickramasuriya et al., 2019) measures the number of nights Australians spent away from home. It includes 228 monthly observations of Visitor Nights (VN) from January 1998 to December 2016, and has a cross-sectional grouped structure based on a geographic hierarchy crossed by purpose of travel. The geographic hierarchy comprises seven states, 27 zones, and 76 regions, for a total of 111 nested geographic divisions. Six of these zones (see Table E.14 in Appendix E) are each formed by a single region, resulting in a total of 105 unique nodes in the hierarchy. The purpose of travel comprises four categories: holiday, visiting friends and relatives, business, and other. To avoid redundancies (Di Fonzo and Girolimetto, 2022b), 24 nodes are not considered, resulting in an unbalanced hierarchy of 525 unique nodes instead of the theoretical 555 with duplicated nodes. The dataset includes the 304 bottom series, which are aggregated into 221 upper time series. Table 5 omits duplicated entries and updates the information in Table 7 from Wickramasuriya et al. (2019). This data can be temporally aggregated into 2, 3, 4, 6, or 12 months ($\mathcal{K} = \{12, 4, 3, 2, 1\}$).

We perform a rolling forecast experiment with an expanding window. The process begins by using the first 10 years, from January 1998 to December 2008, to generate forecasts for the entire following year (2009). Then, the training set is increased by one month. This

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

***** 6 Zones with only one Region are included in Regions. GD: Geographic Division; PT: Purpose of Travel.

Table 5: *Grouped time series for the Australian Tourism Demand dataset.*

process is repeated until the last training set is used (January 1998 to December 2015) with a total of 85 different test sets. For the temporal aggregation dimension we aggregate the monthly data up to annual data. We obtain 12-step, 6-step, 4-step, 3-step, 2-step and 1-step ahead base forecasts respectively from the monthly data and the aggregation over 2, 3, 4, 6, and 12 months. ETS models selected by minimizing the AICc criterion (Hyndman et al., 2023) are fitted to the log-transformed data, with the resulting base forecasts being back-transformed to produce non-negative forecasts (Wickramasuriya et al., 2020).

The (1000) base forecast samples are obtained using the Gaussian approach with sample[8] covariance matrices (Section 4) using multi-step residuals[9] and the bootstrap approach (Section 3.2). For reconciliation, 11 different approaches have been adopted (see Table 2): ct(bu), ct(shr_{cs}, bu_{te}), ct($wlsv_{te}, bu_{cs}$), ct(ols), oct($struc$), oct($wlsv$), oct($bdshr$), oct$_h$($hbshr$), oct$_h$($bshr$), oct$_h$($hshr$), and oct$_h$(shr).

Negative forecasts may be produced during the reconciliation phase (Wickramasuriya et al., 2020; Di Fonzo and Girolimetto, 2022b, 2023b) thus generating unreasonable values (e.g., a negative forecast for tourism demand makes no sense). To overcome this limitation (see Appendix E.1), we applied the simple heuristic proposed by Di Fonzo and Girolimetto (2022a, 2023b). Following Theorem 3.1, we are thus able to obtain reconciled samples respecting non-negativity constraints starting from an incoherent sample simulated from a Gaussian distribution. Finally, to evaluate the performance, we employ the Continuous Ranked Probability Score (CRPS), the Energy Score (ES), and the "Multiple Comparison with the Best" (MCB) Nemenyi test, introduced in Sections 5 and 5.1.

6.1 Results

The CRPS and ES indices are shown, respectively, in Tables 6 and 7 for monthly, quarterly and annual forecasts[10]. 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 bootstrap method is used as a benchmark to calculate the accuracy indices.

Base forecasts using a Gaussian approach are better in terms of both CRPS and ES

[8]The results with shrunk covariance matrices are available in Appendix E.2.

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

[10]The complete results for all temporal aggregation levels are reported in Appendix E.2.

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

Column group header: **Generation of the base forecasts sample paths**

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

Table 6: *$\overline{RelCRPS}$ defined in* (10) *and* (11) *for the Australian Tourism Demand dataset. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

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

In the Gaussian case, bottom-up ct(bu) and partly bottom-up techniques like ct(shr_{cs}, bu_{te}) and ct($wlsv_{te}, bu_{cs}$) lead to better results than the benchmark (bootstrap base forecasts). However, it's not always guaranteed that the improvement is higher than the starting base

Reconciliation approach	Generation of the base forecasts sample paths									
	ctjb	Gaussian approach*				ctjb	Gaussian approach*			
		G	B	H	HB		G	B	H	HB
	$\forall k \in \{12, 6, 4, 3, 2, 1\}$					$k = 1$				
base	1.000	0.956	0.955	0.958	0.951	1.000	0.952	0.950	0.952	0.950
ct(bu)	2.427	0.983	0.983	0.983	0.983	1.759	0.982	0.982	0.982	0.982
ct(shr_{cs}, bu_{te})	1.243	0.886	0.879	0.886	0.879	1.098	0.929	0.928	0.930	0.927
ct($wlsv_{te}, bu_{cs}$)	1.499	0.977	0.977	0.971	0.972	1.241	0.975	0.975	0.973	0.974
oct(ols)	0.955	0.893	0.891	0.893	0.888	0.975	0.937	0.936	0.936	0.935
oct($struc$)	1.085	0.917	0.915	0.916	0.912	1.027	0.943	0.942	0.943	0.942
oct($wlsv$)	1.132	0.933	0.929	0.931	0.927	1.050	0.951	0.949	0.950	0.949
oct($bdshr$)	1.047	0.904	0.897	0.897	0.891	1.009	0.936	0.933	0.934	0.931
oct$_h$($hbshr$)	0.956	0.889	0.886	0.888	0.884	0.975	0.937	0.936	0.937	0.935
oct$_h$($bshr$)	**0.931**	**0.867**	**0.866**	**0.863**	**0.860**	**0.965**	**0.927**	0.927	0.925	0.923
oct$_h$($hshr$)	1.081	0.935	0.931	0.935	0.927	1.028	0.952	0.951	0.952	0.950
oct$_h$(shr)	1.068	0.899	0.878	0.875	0.864	1.023	0.935	**0.923**	**0.921**	**0.916**
	$k = 3$					$k = 12$				
base	1.000	0.961	0.958	0.960	0.955	1.000	0.942	0.947	0.951	0.937
ct(bu)	2.428	0.998	0.997	0.997	0.997	2.990	0.922	0.921	0.923	0.923
ct(shr_{cs}, bu_{te})	1.245	0.911	0.904	0.911	0.904	1.326	0.779	0.767	0.777	0.766
ct($wlsv_{te}, bu_{cs}$)	1.500	0.991	0.991	0.986	0.987	1.679	0.917	0.917	0.906	0.908
oct(ols)	0.976	0.918	0.915	0.917	0.912	0.872	0.783	0.784	0.783	0.779
oct($struc$)	1.096	0.939	0.936	0.938	0.933	1.077	0.826	0.822	0.823	0.818
oct($wlsv$)	1.142	0.953	0.949	0.951	0.946	1.149	0.851	0.845	0.847	0.840
oct($bdshr$)	1.060	0.926	0.920	0.921	0.915	1.021	0.808	0.796	0.796	0.787
oct$_h$($hbshr$)	0.975	0.915	0.912	0.915	0.909	0.872	0.775	0.772	0.772	0.770
oct$_h$($bshr$)	**0.954**	**0.895**	**0.895**	**0.892**	**0.887**	**0.833**	**0.741**	**0.741**	**0.737**	**0.735**
oct$_h$($hshr$)	1.093	0.955	0.951	0.956	0.949	1.066	0.851	0.846	0.848	0.838
oct$_h$(shr)	1.082	0.923	0.903	0.900	0.890	1.043	0.797	0.768	0.764	0.750

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

Table 7: *ES ratio indices defined in* (10) *and* (11) *for the Australian Tourism Demand dataset. Approaches performing worse than the benchmark (bootstrap base forecasts, ctjb) are highlighted in red, the best for each column is marked in bold, and the overall lowest value is highlighted in blue. The reconciliation approaches are described in Table 2.*

forecasts (by comparing the value of each column). This is particularly true for higher levels of temporal aggregation (see Appendix E.2 for details). Overall, oct($bdshr$) in terms of CRPS is almost always the best. The shrinkage approach oct$_h$($hshr$) performs well in the bootstrap case: it is competitive with oct($bdshr$) at lower temporal frequency ($k \in \{2, 1\}$) and it is able to improve for $k \geq 3$. In terms of ES, oct($bdshr$) is still competitive, although it does not always show the best relative performance. In this case, approaches that attempt to explicitly take into account temporal and cross-sectional relationships, such as oct$_h$($hbshr$) and oct$_h$($bshr$), perform better. It is also worth noting that techniques that don't make use of residuals like oct(ols) and oct($struc$) prove to be competitive by consistently improving on the base forecasts in terms of both CRPS and ES.

Figure 6 shows the MCB using the CRPS for the Gaussian approach using multi-

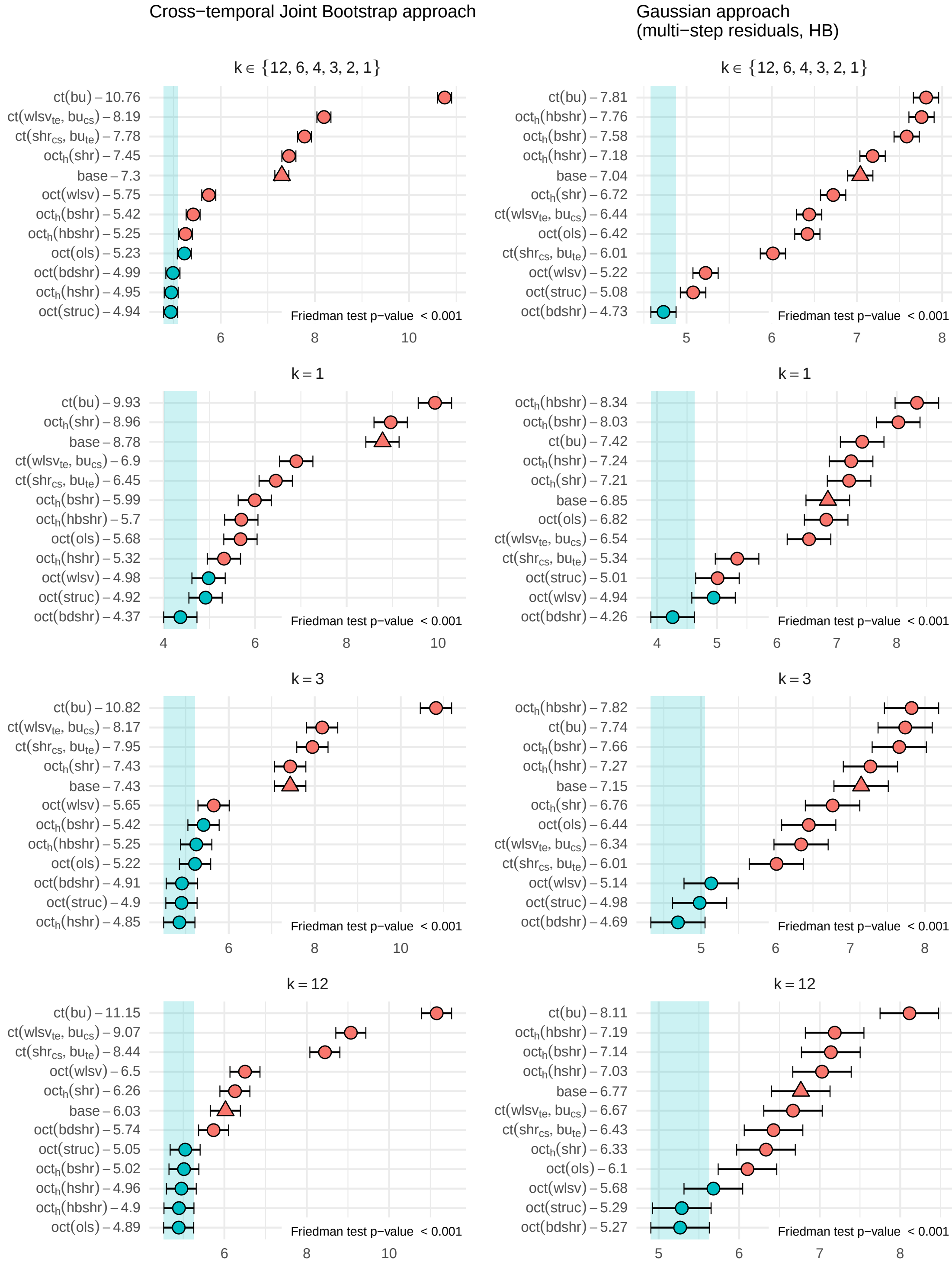

Figure 6: *MCB Nemenyi test for the Australian Tourism Demand dataset using CRPS at different temporal aggregation levels for the Gaussian (multi-step residuals, HB) and the non-parametric bootstrap approaches. In each panel, the Friedman test p-value is reported in the lower right corner. The mean rank of each approach is shown to the right of its name. Statistical differences in performance are indicated if the intervals of two forecast reconciliation procedures do not overlap. Thus, approaches that do not overlap with the blue interval are considered significantly worse than the best, and vice-versa.*

step residuals (HB) and the non-parametric bootstrap approach. In general, the partly bottom-up procedure improves with respect to base forecasts at monthly level, but optimal cross-temporal procedures are always better. In the bootstrap framework, we can identify a group of three procedures, oct($bdshr$), oct($hshr$) and oct($struc$) that are almost always in the group of the best approaches (denoted by the blue dot). In the Gaussian framework, oct($wlsv$), oct($struc$), and oct($bdshr$) are always significantly better than base forecasts and equivalent in terms of results for temporal aggregation orders greater than 2. For monthly series, oct($bdshr$) is always significantly better than all other approaches.

7 Conclusion

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

Moreover, we analyze the use of residuals, showing that one-step residuals fail to capture the temporal structure, and propose multi-step residuals that can fully capture the cross-temporal relationships investigating them in a simulation (Appendix C). Due to the high-dimensionality of the cross-temporal setting when dealing with covariance matrices, we propose four alternative forms to reduce the number of parameters to be estimated, showing that the overlapping residuals may reduce the high-dimensionality burden by increasing the number of available residuals. These ideas are worth requiring further investigation in future works.

Finally, we perform empirical applications on two datasets commonly used in forecast reconciliation research: Australian GDP from Income and Expenditure sides and Australian Tourism Demand. We find that in both cases optimal cross-temporal reconciliation approaches significantly improve on base forecasts. We also compare these with partly bottom-up techniques that use uni-dimensional reconciliation (either cross-sectional or temporal) and confirm that simultaneously exploiting both dimensions in reconciliation produces better results, especially at higher levels of temporal aggregation.

In conclusion, our findings confirms that reconciliation approaches are an important tool in the field of operational research to improve the accuracy of forecasts while simultaneously ensuring their coherency both in space and time. Furthermore, these techniques can also be customized to suit the specific needs of an organization, allowing for the incorporation of relevant domain-specific knowledge (e.g., non negative constraints) and expertise, ensuring that the resulting forecasts are not only accurate but also coherent and more reliable for decision-making purposes.

Acknowledgments

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

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

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

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

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

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

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

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

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

Di Fonzo, T. and D. Girolimetto (2022b). Forecast combination-based forecast reconciliation: Insights and extensions. *International Journal of Forecasting*. in press.

Di Fonzo, T. and D. Girolimetto (2022c). Point and probabilistic forecast reconciliation for general linearly constrained multiple time series. *Working paper*.

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

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

Dunn, D. M., W. H. Williams, and T. L. Dechaine (1976). Aggregate versus Subaggregate Models in Local Area Forecasting. *Journal of the American Statistical Association 71*(353), 68–71.

Eckert, F., R. J. Hyndman, and A. Panagiotelis (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. and C. Morris (1975). Data Analysis Using Stein's Estimator and its Generalizations. *Journal of the American Statistical Association 70*(350), 311–319.

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

Fleming, P. J. and J. J. Wallace (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. and T. Di Fonzo (2023). *FoReco: Point Forecast Reconciliation.* R package v0.2.6.

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

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

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

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

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

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

Kourentzes, N. (2022). Toward a one-number forecast: Cross-temporal hierarchies. *Foresight: The International Journal of Applied Forecasting 67*, 32–38.

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

Kourentzes, N. and G. Athanasopoulos (2021). Elucidate structure in intermittent demand series. *European Journal of Operational Research 288*(1), 141–152.

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

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

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

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

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

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

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

Papadakis, M., M. Tsagris, M. Dimitriadis, S. Fafalios, I. Tsamardinos, M. Fasiolo, G. Borboudakis, J. Burkardt, C. Zou, K. Lakiotaki, and C. Chatzipantsiou (2022). *Rfast: A Collection of Efficient and Extremely Fast R Functions*. R package version 2.0.6.

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

Punia, S., S. P. Singh, and J. K. Madaan (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*. Vienna, Austria: R Foundation for Statistical Computing.

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

Schäfer, J. and K. Strimmer (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., F. Petropoulos, N. Kourentzes, and V. Assimakopoulos (2020). Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Applied Energy 261*, 114339.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.

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

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

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

Yagli, G. M., D. Yang, and D. Srinivasan (2020). Reconciling solar forecasts: Probabilistic forecasting with homoscedastic Gaussian errors on a geographical hierarchy. *Solar Energy 210*, 59–67.

Yang, D. (2020). Reconciling solar forecasts: Probabilistic forecast reconciliation in a nonparametric framework. *Solar Energy 210*, 49–58.

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